

# The ION GENERATOR, a Heuristic Program in Organic Mass Spectrometry

(Un programme heuristique pour la Spectrométrie de masse,  
connu sous le nom de ION GENERATOR)

THESE

présentée à la Faculté des sciences de l'Université de Genève  
pour obtenir le grade de docteur ès sciences chimiques

par

Allan B. DELFINO

de

SAN FRANCISCO (Californie)

Thèse N° 1675

GENEVE

Editions Médecine et Hygiène

1974



Digitized by the Internet Archive  
in 2026

# The ION GENERATOR, a Heuristic Program in Organic Mass Spectrometry

(Un programme heuristique pour la Spectrométrie de masse,  
connu sous le nom de ION GENERATOR)

THESE

présentée à la Faculté des sciences de l'Université de Genève  
pour obtenir le grade de docteur ès sciences chimiques

par

Allan B. DELFINO

de

SAN FRANCISCO (Californie)

Thèse N° 1675

GENEVE  
Editions Médecine et Hygiène  
1974

La Faculté des sciences, sur le préavis du professeur ordinaire A. Buchs, directeur de thèse et d'une commission composée du professeur ordinaire B. Levrat (Faculté des sciences) et du professeur R. Tabacchi (Université de Neuchâtel), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 12 juillet 1974

*Le doyen :*  
Jean Muller

Thèse No 1675



Professor Armand Buchs, Director of the Mass Spectrometry Laboratory of the University of Geneva, expressed his utmost confidence in my ability to finish this work. As my thesis director, he constantly encouraged me and helped me to finish this work on such a positive note. I sincerely appreciate his friendship and guidance. His time and efforts spent in my behalf will remain to me a constant reminder of how one should act toward others in order to be respected and esteemed.

My earnestly expressed thanks go to Professor Bernard Levrat, member of my thesis committee, who, during the three years that I served as a teaching assistant at the Computer Science Department, constantly encouraged my development in Computer Science, as well as in French.

To Professor Raffaele Tabacchi, member of my thesis committee, I extend my heartfelt thanks for his time spent in my behalf.

My sincere appreciation is acknowledged for the fine computing services maintained by the operations personnel of the Centre Cantonal d'Informatique. All of the programming done for this work was done on the CDC 3800 which functions under their guidance.

Finally, I would like to thank my co-workers in the Mass Spectrometry Laboratory, in particular, Mr. Werner Kloeti, Mrs. Françoise Kloeti, Miss Brigitte May and Mr. Paul Hugentobler, for their assistance and friendship.

To

ANNE

*without whose encouragement, cheerfulness and patience, this work would not have been possible.*

MY PARENTS

*who have always approved of my plans toward education, no matter what would be the cost to themselves.*

T A B L E O F C O N T E N T S

|                                                                                    | <u>Page</u> |
|------------------------------------------------------------------------------------|-------------|
| I. INTRODUCTION                                                                    | 1           |
| II. SURVEY OF COMPUTER PROGRAMS IN ORGANIC MASS SPECTROMETRY                       | 2           |
| II.1 Overview                                                                      | 2           |
| II.2 Programs which study the fragmentation of known compounds                     | 2           |
| II.3 Programs solving "unknowns"                                                   | 3           |
| II.4 Programs predicting mass spectra                                              | 4           |
| III. BACKGROUNDS IN HEURISTIC COMPUTER PROGRAMMING AND ELECTRON-BOOKKEEPING THEORY | 6           |
| III.1 Heuristic Computer Programming                                               | 6           |
| III.1.1 The Chemical Heuristic                                                     | 9           |
| III.2 The electron-bookkeeping theory                                              | 9           |
| IV. PRIMITIVE REACTION STEPS OF THE ION GENERATOR PROGRAM                          | 13          |
| IV.1 Introduction                                                                  | 13          |
| IV.2 Weak bonds and weakly-held hydrogen atoms                                     | 13          |
| IV.3 Stable environments for radicals and charges                                  | 14          |
| IV.4 Ring and transition state flexibilities                                       | 16          |

|         |                                                                                    |    |
|---------|------------------------------------------------------------------------------------|----|
| IV.5    | Primitive reactions: their motivating conditions and resultant transforms          | 17 |
| IV.5.1  | Ionization of the molecule                                                         | 18 |
| IV.5.2  | Simple homolysis ( $\alpha$ -cleavage )                                            | 18 |
| IV.5.3  | Heteroatom heterolysis                                                             | 19 |
| IV.5.4  | Simple hydrogen rearrangement (cyclic transition state)                            | 20 |
| IV.5.5  | Formation of a bond from two radicals                                              | 20 |
| IV.5.6  | Radical site motivated ring closure                                                | 21 |
| IV.5.7  | One-electron bond dismantling                                                      | 21 |
| IV.5.8  | Non-heteroatom heterolysis and skeletal rearrangement                              | 22 |
| IV.5.9  | Homolysis with satellite rearrangement                                             | 23 |
| IV.5.10 | Heterolysis of a two atom functional group                                         | 24 |
| IV.5.11 | Retro-Diels-Alder                                                                  | 24 |
| IV.5.12 | Hydrogen Push Rearrangement                                                        | 24 |
| IV.5.13 | Double bond triggered hydrogen rearrangement, step I (McLafferty rearrangement)    | 25 |
| IV.5.14 | Double bond triggered hydrogen rearrangement, step II (McLafferty rearrangement)   | 26 |
| IV.5.15 | Homolytic overbond relief of a heteroatom                                          | 26 |
| IV.5.16 | Construct a one-electron bond from adjacent charge and radical sites (stop action) | 27 |
| IV.5.17 | Bond replacement by a hydrogen atom bond to an overbonded heteroatom               | 27 |
| V.      | HOW TO USE THE ION GENERATOR PROGRAM                                               | 29 |
| V.1     | General guidelines                                                                 | 29 |
| V.2     | CDC 3800 control cards                                                             | 29 |
| V.3     | Prologue data set                                                                  | 30 |
| V.3.1   | Common set-up options                                                              | 30 |
| V.3.2   | Periodic table input                                                               | 33 |

|             |                                                                                                                         |    |
|-------------|-------------------------------------------------------------------------------------------------------------------------|----|
| V.3.3       | Ionization table input                                                                                                  | 34 |
| V.4         | Structure-spectra data set(s)                                                                                           | 36 |
| V.4.1       | Identification card                                                                                                     | 36 |
| V.4.2       | The comparison spectrum                                                                                                 | 36 |
| V.4.3       | Structure input                                                                                                         | 37 |
| V.5         | Output of the ION GENERATOR program                                                                                     | 38 |
| V.5.1       | The mass-only mass spectrum                                                                                             | 38 |
| V.5.2       | Mechanism reports                                                                                                       | 39 |
| V.5.3       | Linear notation for structural outputs                                                                                  | 40 |
| VI.         | RESULTS AND DISCUSSION                                                                                                  | 43 |
| VI.1        | The early efforts                                                                                                       | 43 |
| VI.2        | The change in strategy                                                                                                  | 44 |
| VI.3        | Recent studies with the ION GENERATOR working with cyclic compounds                                                     | 46 |
| VI.3.1      | Reactions motivated by one-electron bonds                                                                               | 47 |
| VI.3.2      | Reactions involving six-membered rings                                                                                  | 48 |
| VI.3.3      | The effect of double bond conjugation                                                                                   | 49 |
| VI.3.4      | Ionization of resonance rings                                                                                           | 51 |
| VI.4        | Possibilities for future research                                                                                       | 52 |
| VII.        | BIBLIOGRAPHY                                                                                                            | 53 |
| Appendix A. | Mass Spectra and Computers                                                                                              | 60 |
| Appendix B. | Heuristic Programming as an Ion Generator in Mass Spectrometry. I. Generation of Primary Ions with Charge Localization. | 91 |

|             |                                                                                                                               |     |
|-------------|-------------------------------------------------------------------------------------------------------------------------------|-----|
| Appendix C. | Heuristic Programming as an Ion Generator in Mass Spectrometry. II. Generation of Primary Ions from Simple Cyclic Structures. | 106 |
| Appendix D. | Heuristic Programming Techniques Applied to Chemical Mass Spectrometry.                                                       | 118 |
| Appendix E. | Heuristic Programming as an Ion Generator in Mass Spectrometry. III. Fragmentation of Aliphatic Compounds                     | 126 |
| Appendix F. | PROGRAM DESCRIPTIONS OF THE ION GENERATOR SYSTEM                                                                              | 139 |
| F.1         | Introduction                                                                                                                  | 139 |
| F.2         | ION GENERATOR, the main program                                                                                               | 140 |
| F.3         | Overlay I, Prologue data set input                                                                                            | 140 |
| F.3.1       | Programs of overlay I                                                                                                         | 140 |
| F.4         | Overlay II, the mechanism generator                                                                                           | 141 |
| F.4.1       | Programs of overlay II                                                                                                        | 142 |
| F.5         | Overlay III, the mechanism sort and print                                                                                     | 150 |
| F.5.1       | Segment II, overlay III, program descriptions                                                                                 | 150 |
|             | Résumé en français                                                                                                            | 153 |

## Chapter I.

### I N T R O D U C T I O N

In recent years, the digital computer has had an increasing impact on life in general. Science has benefited greatly from the advances made in computer technology because of the tremendous calculating speeds which are currently achieved. More complex problems are being solved with each increment of speed that becomes available. At the same time, advances made in programming languages, such as FORTRAN /1/ and LISP /2/, have made it possible for the scientist to write his own computer programs, instead of having to depend upon a computer specialist. These improvements and developments have taken place during the last three decades, but it is only in the last five years that we have seen an impact on chemistry by non-calculational programs.

A non-calculational program, in the broadest sense, is a program whose results are not solely numerical values. The results of this type of computer program is a logical construct of some sort: a molecular structure, a method for synthesizing compounds, or a mass spectral fragmentation mechanism, just to name a few. An example of a calculational program would be a molecular orbital calculation in quantum chemistry. We are not speaking of a difference in complexity, but really of a difference in kind.

Two important programs, which result in constructs, are those of Corey et al /3 - 6/ and Bersohn /7/. Their work outlines routes of synthesis to a given target structure. The input is the structure, for which a synthesis is sought. The output is a series of known synthetic steps which will transform an available starting material to the target structure. Of course, these programs do not invent chemistry, i.e., propose new synthetic steps. On the contrary, they are given a "book" of chemical reactions that have been discovered by chemists; from these reactions complete syntheses are proposed.

We have, in the field of mass spectrometry, many examples of non-calculational programs. In Chapter II we will review the different programming areas of mass spectrometry; in later chapters, we will demonstrate why the ION GENERATOR (IG) program, the subject of this thesis, is an important contribution to mass spectrometry.

## Chapter II.

# SURVEY OF COMPUTER PROGRAMS IN ORGANIC MASS SPECTROMETRY

### II.1 Overview

Low-resolution organic mass spectrometry has drawn an enormous amount of attention from teams of chemists and computer specialists. The results have been at once quite varied and quite encouraging. In order to put the work of this thesis into a useful perspective, we will limit the discussion to non-calculational programs. The work done in the field of data-acquisition, monitoring and file searching relating to mass spectrometry is mentioned in the review paper by Delfino and Buchs / *Appendix A* /; it will not be discussed here.

For the purposes of clarity, the field of mass spectral computer programs has been arbitrarily divided into three parts: programs which study fragmentations of known compounds, programs which solve unknown structures from their mass spectra, and programs which predict mass spectra. More detailed descriptions of the programs to be mentioned below are given in *Appendix A*.

### II.2 Programs Which Study the Fragmentation of Known Compounds

When a molecule is bombarded with 70 eV. electrons, a reproducible set of ions is generated. It is this reproducibility that has rendered mass spectrometry useful as an analytical tool. But the resultant ions are not always simple structural sub-units of the parent structure. This fact necessitated an enormous effort in the study and elucidation of fragmentation mechanisms. The overall goal has been a general understanding of the step-by-step processes underlying mass spectral fragmentations for each chemical class, and if possible, for all classes, i.e., a unified theory of organic mass spectrometry.

Since the fifties, thousands of compounds have been submitted to fragmentation studies, generally with the help of labelled (mostly  $^2\text{H}$ )

analogs. This has resulted in a pool of partially classified knowledge, from which an attempt has been made to develop a theory. The DENDRAL group of Djerassi, Feigenbaum and Lederberg and their co-workers is developing a program which will draw *itself* theoretical inferences and deductions from a set of related compounds. This program, named METADENDRAL, is briefly described in Appendix A; it has been used recently to study estrogenic steroids /8 - 12/.

### II.3 Programs Solving "Unknowns"

Mass spectrometry can solve the structure of an unknown molecule from just its mass spectrum, or little else, in many cases. In order to do routine structural analyses, many non-calculational programs have been written. The motivation has been the rapid, and possibly more accurate, processing of large numbers of spectra.

An outstanding example of such a program has been that of Delfino et al /13/. This program, described in Appendix A, with just the low-resolution mass spectrum, can deduce the correct structure for any compound of the class of saturated, acyclic, monofunctional substances; this includes amines, alcohols, ethers, thiols and thioethers. Although it does not always deduce just a single structure, the correct one is always among the output structures.

Another sort of non-calculational program, which was developed for structure determinations, assembles structural features that are postulated as a result of finding certain diagnostic ion series in the spectrum. Venkataraghavan et al /A:37 (the thirty-seventh reference in Appendix A), 14 - 16/ and Crawford and Morrison /A:34/ have developed programs which they claim demonstrate promise for subsequent generalizations. Complete descriptions of these programs are to be found in Appendix A.

Jurs et al /A:15 - A:23, 17, 18/ pioneered the use of binary pattern classifiers to solve structures in mass spectrometry. A binary classifier selects one of two categories for spectra, by means of vectors which are multiplied with the spectrum. A series of such vectors can do very complex classifications. The early work of binary pattern classifications was reviewed in Appendix A. Since Appendix A has appeared, numerous workers have published results about pattern classifying programs. Isenhour and his co-workers /19, 20/ are concentrating on discriminant functions and the determination of the information content of mass spectra. Theirs is an approach to the understanding of the limitations of pattern recognition. Clerc and Seibel/ 21 / have also

pointed out that improper conclusions can be drawn from pattern classification. Bender and Kowalski and various co-workers /22 - 24/ are examining the effects of various transformations on the classification process. Stonham et al /25/ have implemented a totally electronic, binary pattern classifier, which constitutes a novel approach. Numerous other articles by various authors /26 - 45/ have been published very recently in the general area of the determination of unknown structures by mass spectrometry with the aid of digital computers.

#### *II.4 Programs Predicting Mass Spectra*

The subject of this thesis, the ION GENERATOR program, when given any molecule, proposes a series of ionic structures. This, in effect, can be considered the prediction of a mass-only mass spectrum. A low-resolution mass spectrum is a collection of various unit masses, each occurring with an intensity which is related to the relative abundance that that ion had in relation to the strongest (base) peak. A mass-only mass spectrum is, on the other hand, a spectrum for which no relative intensities are assigned to the generated ions. Before launching into a detailed description of how the ION GENERATOR program operates when it predicts a mass spectrum, let us examine briefly the other approaches that have been tried for the digital computer prediction of mass spectra.

The quasi-equilibrium theory (QET) proposed by Rosenstock /46/ and clearly described by Kiser /47/, is a method based on the absolute rate theory of unimolecular reactions. It requires that the rate constants of all of the unimolecular reactions, occurring to give rise to the mass spectrum, be calculated. Even for molecules that are not really complex, the specification of all of the unimolecular decomposition reactions is difficult, if not impossible. Without these, one cannot hope to calculate the intensities of the mass spectrum via the quasi-equilibrium theory. Where the IG program specifies reactions resulting in structures (not necessarily the complete set), but says nothing about intensities, the QET must know the complete set of reactions, and then with certain assumptions, can calculate the intensities.

As will be shown in later chapters, the IG program is a general program never needing to consider the chemical class of the structure upon which it is working. There have been two programs which have been written for restricted classes of compounds. First, a program described by Djerassi et

al /A:41, A:42 /, called the PREDICTOR, predicts a crude mass spectrum for the restricted classes: aliphatic ketones, ethers and amines. Secondly, Jurs and Schecter /48, 49/ have applied the techniques of learning machines (classifiers), to predict the presence or absence of masses on the low-resolution mass scale. This has been achieved for molecules restricted to the molecular formula range  $C_{3 \rightarrow 10}H_{2 \rightarrow 22}O_{0 \rightarrow 4}N_{0 \rightarrow 2}$  and a limited set of 61 structural sub-units.

Presently, it can be said that the IG program is the only computer program which is completely general. It has no arbitrary restrictions, except that the structure must have 50 or less non-hydrogen atoms (a size-of-computer restriction); they can be connected in any way whatsoever. The program at no time tries to identify the compound by class or by functionality.

## Chapter III.

### BACKGROUNDS IN HEURISTIC COMPUTER PROGRAMMING AND ELECTRON-BOOKKEEPING THEORY

Organic mass spectrometry deals with the energy-induced fragmentation of organic compounds (usually via electron impact). Compounds give rise to ions whose masses are analyzed and whose concentrations in the ion beam are measured. The ION GENERATOR program has been designed to effect 70 eV. fragmentations, based on a theory which has evolved from the experimental observation of the relationships which exist between structures and their breakdown patterns. All these observations have been formalized in treatises on organic mass spectrometry, such as Djerassi et al /50 - 52/. This formalism has permitted a coherent understanding of many of the mechanisms which might be occurring in the mass spectrometer, but was not, until now, cast in a form which would permit prediction of ionic structures by heuristic search methods.

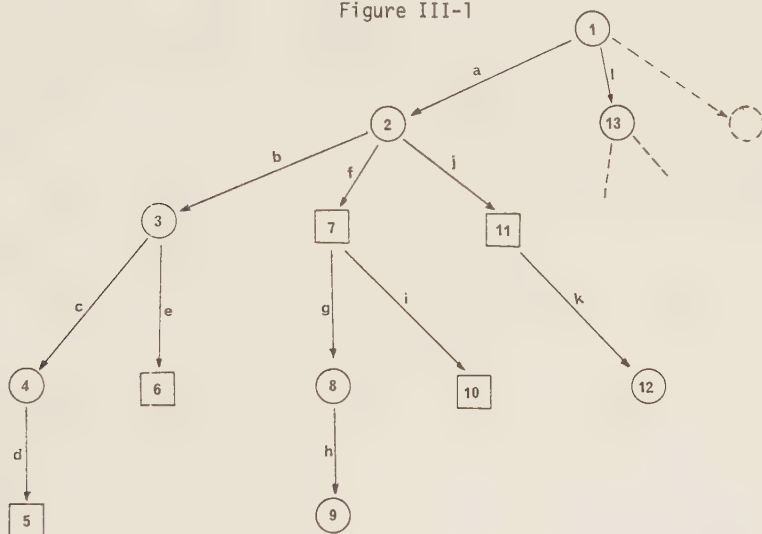
#### III.1 *Heuristic Computer Programming*

Slagle /53/ defines a *heuristic* as "...a rule of thumb, strategy, method, or trick used to improve the efficiency of a system which tries to discover the solutions of complex problems". *Heuristic programming* is then the programming of a digital computer with an algorithm than uses *heuristics*. One common procedure, in such programs, is the search of a complete, irredundant solution space. The heuristics are installed in such a way as to reduce this search of the solution space to the portions of it which are most likely to yield interesting results.

In the case of the IG program, the solution space is a space of ionic structures, which are transforms of the original molecular structure given to the program at the start. These transforms are made by the chemical reactions known to the program, i.e., the theory; this set of reactions is called the set of primitive reactions, or primitives, for short. The solution space is generated by applying all of the primitives in all possible, but chemically plausible, ways, until one of the various heuristics stops the process. The program tries to transform every structure and every descendent structure, until it can go no further.

This kind of process can be characterized by an "inverted tree", where, in the case of the IG program, the trunk is the molecular structure and the branches are the primitives that are transforming a structure.

Figure III-1



In figure III-1, for example, the hypothetical molecule, represented by node No. 1, is transformed to molecular ions (node Nos. 2, 13 and others not shown) by the process of ionization. Ionization to molecular ion No. 2 is done right away; the way to do No. 13 is remembered for use later on. The nodes have been numbered in the order that these structures have appeared. The transforming primitives have been lettered with small letters in the order that the transformations are applied. Hence the second molecular ion is the thirteenth structure, created by transforming structure No. 1 with primitive l. The nodes depicted as squares are structures that appeared as a result of a loss of a neutral species (fragmentation), whereas the nodes depicted as circles are structures resulting from a transformation where no neutral species was lost. No loss occurred when structure No. 2, on which three primitives were found to be possible, was transformed to structure No. 3. Structure No. 3 could support two transformations, the first of which results in structure No. 4, which in turn could sustain only one reaction. This single reaction results in structure No. 5, the first fragmentation to be encountered by the program, i.e., this is the first time that an ionic structure occurred with a mass less than that of the molecular ion. When the expulsion of a neutral species occurs, the

program notes the mechanism (a,b,c,d), the ionic and neutral structures, and their masses.

The program finds structure No. 5 to be a dead-end, requiring that the inverse of reaction d be done to structure No. 5 which in effect reattaches the expelled neutral species, which reconstitutes structure No. 4. The program notes at that point that no other reactions are waiting to be done to structure No. 4, so the inverse of reaction c is done next, returning structure No. 3 for a subsequent reaction; reaction e creates structure No. 6, the second fragment ion encountered in the search. The mechanism for ionic structure No. 6 is recorded as (a,b,e). The other mechanisms presented in Fig. III-1 are:

(a,f) resulting in structure No. 7  
(a,f,i) resulting in structure No. 10  
and (a,j) resulting in structure No. 11.

Structures Nos. 9 and 12 were dead-ends which were not fragmentation nodes. Structure No. 13 is the start of a sub-tree, which is not shown.

An important point should be stressed at this point. If, for example, mechanism (a,b,c,d) happens to be a widely accepted mechanism for the class of compounds, to which the given molecule belongs, this fact remains unknown by the program. The program does this heuristic search of the ionic structure space by constructing mechanisms from all chemically plausible combinations of primitives. At no time does the program make an attempt to determine the chemical class of the given substance, in order to "look-up" the correct mechanism. As is shown later, in Chapter IV, primitive reactions consider only the immediate environment of radical sites, charge sites, double bonds and weak bonds in the structures. Distinction is made, as far as atoms are concerned, only between atoms with lone-pairs of electrons (Lewis bases) and others, including carbon. As far as the ION GENERATOR program is concerned, there are two types of atoms and the various bonding configurations in which they can be.

The realization of the heuristic search, and all of the important programming details of the IG program, are given in Appendix D. The facts contained therein will not be discussed in greater detail at this time.

### III.1.1 The Chemical Heuristic

All possible combinations of the 16 primitives could be generated, no matter how improbable they might seem from a chemical point of view. The disqualifying conditions that are examined before a primitive reaction step is declared possible, on a subject structure, for inclusion in a combination, are the *chemical heuristics* of the IG program. They reduce the number of uninteresting mechanisms which would be proposed were they not included. These heuristics are not energy or bond strength calculations, but are simply *chemical notions*, like those chemists are prone to use when discussing chemistry at the blackboard, expressed as computer programs. These notions include: stability of free radicals and charge centers, ring flexibility, limited hydrogen rearrangements and the availability of double bonds, to name just a few.

The IG program represents the first time that the concept of chemical reasoning has ever been used as the basis of a predictive theory. Most theoretical programs make calculations based on a mathematical model to arrive at the desired quantities. But the IG program arrives at the structures of a mass spectrum by working with a symbolic model of the molecule, manipulating it as would the chemist at the blackboard. The most interesting feature of this theory is the simplicity of the various considerations that are made. This simplicity, coupled with the degree of agreement that the predicted spectra have with the actual recorded spectra, is indeed a concrete demonstration of the power of the chemical reasoning process.

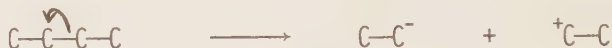
### III.2 The Electron-Bookkeeping Theory

Electron-bookkeeping, a term coined by Djerassi et al /50 - 52/, implies that discussions of reactions, occurring in the mass spectrometer, should take proper account of electrons in any transformation. It is this chemical reasoning process, or theory, upon which the IG program is based.

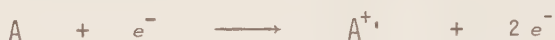
There are two bonding transformations in electron-bookkeeping. First, there is homolytic fission, denoted by a half-arrow, the so-called "fish-hook symbol".



Secondly, the heterolytic cleavage, which involves two electrons, is denoted by the full arrow.

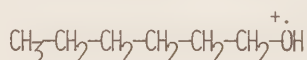


These processes are employed to rationalize the chemistry occurring inside the ion source of the mass spectrometer, where electron bombardment gives rise to ions by the following reaction.

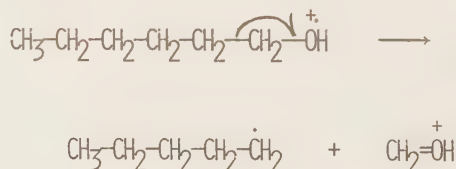


The dual presence of a charge site and a radical site are the main motivating factors which trigger heterolytic or homolytic cleavages respectively.

Mass spectrometrists explain the formation of ions in terms of homolytic and heterolytic cleavages, after they have decided, on chemically plausible grounds, that the charge should be localized at a particular site in the molecule. For example, in hexane-1-ol, they localize the charge on the oxygen atom because they agree that ionization affects the lone pair more readily than electrons engaged in  $\sigma$  bonds. This results in the following presentation of the molecular ion of hexane-1-ol.



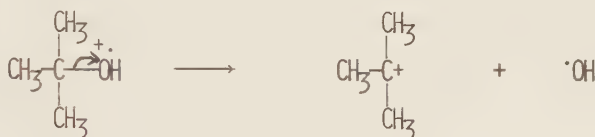
The presence of an unpaired electron on the oxygen atom is then the factor which triggers what is called  $\alpha$ -cleavage, i.e., homolysis of a bond, resulting in the creation of another unpaired electron on an adjacent atom; this allows the energetically favorable process of bond formation to occur. This commonly held notion was supported recently by the molecular orbital calculations of Ichikawa and Ogata /54/.



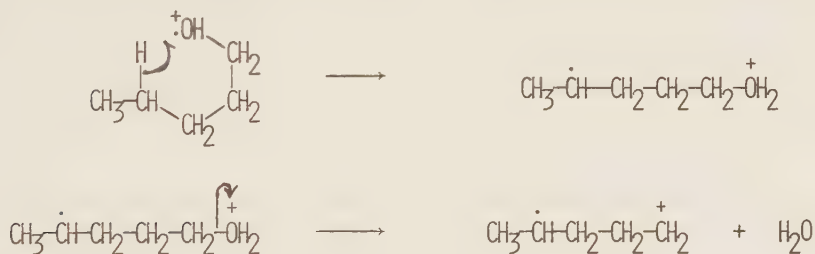
In hexane-1-ol, heterolytic cleavage can also occur.



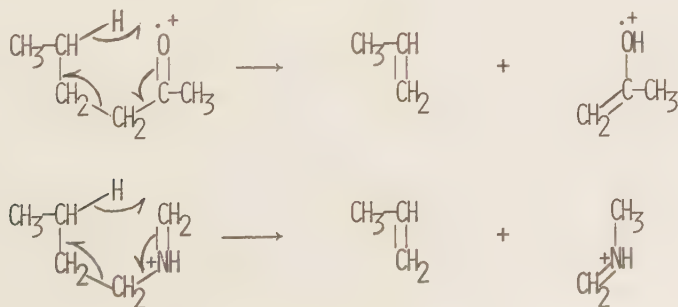
This heterolysis gives rise to a hydrocarbon ion in alcohols. Since substituted carbonium ions are more stable than unsubstituted carbonium ions, such as that which arose due to the above reaction, one would expect that this reaction in a compound, such as tertiary butyl alcohol, would result in a strong ion.



The well-established 1,4 elimination of water in long-chain alcohols is rationalized by mass spectrometrists in terms of a homolysis, followed by a heterolysis.



Some mechanisms are found in more than one class. The McLafferty rearrangement /55/, for example, gives rise to a primary ion in ketones, but to secondary ions in amines having the proper structure.



The differences between the normal electron-bookkeeping theory and the version of it incorporated in the ION GENERATOR program are very great. Mass spectrometrists use these constructs, called mechanisms, to make a *posteriori* explanations (rationalizations) of ionic appearances in mass spectra. The IG's form of this theory is used to make a *priori* predictions of probable ionic fragmentations. Hence, the essential contribution to mass spectrometry, which results from the work done on the IG program, is the specification of the groups of homolyses and heterolyses, called the primitive reactions. It is these primitive reactions that can be applied over and over again in ionic structures as long as the required bond environments exist in the structure. The complete specification of the primitive reactions and their required environments is given in the following chapter, Chapter IV.

## Chapter IV.

### PRIMITIVE REACTION STEPS OF THE ION GENERATOR PROGRAM

#### *IV.1 Introduction*

In section III.2, the concept of a heuristic search among the mechanisms, proposed to predict the mass-only mass spectrum of a structure, was discussed. It was mentioned that mechanisms are combinations of indivisible reaction steps. These are termed "primitive" reaction steps, or just primitives, according to the usage common in heuristic programming /53/. Sixteen primitive reaction steps were found necessary to attain satisfactory reaction schemes. Every one of these sixteen can be found, in fact, in mass spectrometry textbooks /50-52, 56-59/; but never before was an attempt made to enumerate a list of the needed basic chemical transformations generating the greater part of the mechanisms given in these textbooks as examples.

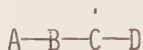
The use of any one of these sixteen primitives must always conform with some motivating conditions; like the chemical reasoning that gave birth to electron-bookkeeping, they are expressions of the probable consequences of structural environments. They also represent syntheses that have been made by numerous mass spectrometrists, after having studied spectra, and having explained the fragmentation modes found in these spectra with the aid of the electron-bookkeeping theory. These syntheses have rarely been supported by theoretical calculation /54/; in most cases, they are widely held notions, supported indirectly by experimental observations. Some additional conditions, or heuristics, were born of the program's need to have a finite amount of output to print and a finite amount of work to do.

#### *IV.2 Weak Bonds and Weakly-Held Hydrogen Atoms*

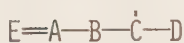
The most probable cleavages to occur will be those which affect mainly the weaker bonds of the structure; the most probable rearrangements involve weakly-held hydrogen atoms, when no particularly favorable transition state ring size can be formed. A sub-program of the IG prepares a table of these weak bonds; this table is accessible to the chemical reaction surveyor programs that

apply weak-bond conditions to their respective primitive reactions. A bond is considered weak when:

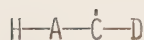
- a) *it is in the  $\alpha$ -position with respect to a radical site.* The annotated examples below illustrate this point.



Bond A-B is weak.



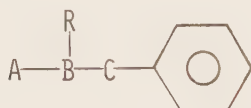
Bond A-B is not weak, because it is a vinylic bond nullifying the effect of the radical site.



The H atom on atom A is weakly-held.

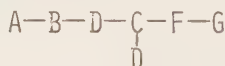
This rule accounts for the tendency of electrons to pair-up in the ionic or neutral species, and for the fact that vinylic cleavages are unfavorable.

- b) *it is in the  $\beta$ -position with respect to a multiple bond.* Allylic bonds tend to rupture more readily than other bonds. An example situation is given below.



If bond B-C is not also vinylic and both B and C are not electron deficient, then it is weak, being allylic to the phenyl ring.

- c) *it is efferent from a secondary or tertiary carbon atom.* Since substituted carbonium ions are more stable than primary carbonium ions, bonds in such environments tend to rupture.



If C is a secondary carbon atom, then bonds C-D, C-E and C-F are considered weak

- d) *it is a bond connecting an H atom of acidic character.* In this case, there must be a strong acceptor site for the hydrogen atom in the molecule.

#### IV.3 Stable Environments for Radicals and Charges

One guiding concept of reaction dynamics is that reactions involving

the transformations of stable species are inhibited. As the internal energy decreases going from molecular ions to daughter ions, all reactions become less probable in daughter ions. Hence the IG is constantly searching the molecule for stable situations and removes them from consideration as potential reactants. One important stable situation, which the program uses in this way, is the stabilized or deactivated radical site.

A radical site is considered by the IG program to be deactivated if it is in any of the following environments:



a



b



c



d

- a) pairing is favored.
- b) resonance stabilized.
- c) pairing to form a one-electron bond is favored.
- d) resonance stabilized.

A radical site which is allylic to a resonance ring, such as a phenyl group, fits into category b.

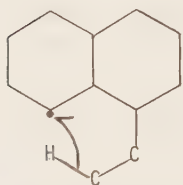
A line of reasoning similar to that of deactivated radical sites is used by the IG when it searches for double bonds to motivate reactions. Double bonds do not trigger reactions when they are conjugated. In every other situation, double bonds cause transformations to occur, even if the double bond was formed by the IG in another reaction.

In keeping with the concept that a decrease in internal energy decreases reaction probabilities, certain reactions use different motivating conditions when their surveyor program recognizes that it is no longer working on the molecular ion. Reaction sequences not leading to fragmentation are controlled somewhat to limit the number of times that the same transformation, with differing reaction step sequences, would happen. Certain reactions are allowed to occur only when there are no more radical sites left in the ionic structure; it is based on the idea that the activation energy for skeletal rearrangements in carbonium ions is much higher than reactions involving radical sites.

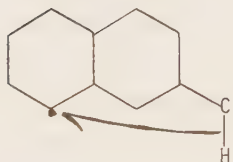
#### IV.4 Ring and Transition State Flexibilities.

In order that chemically improbable rings will not be formed, nor that hydrogen rearrangements of low probability will occur, the IG makes an attempt to identify cycles or cyclic transition states which are strained; i.e., those which would be inflexible. The following statements describe the algorithmic considerations which are made to determine inflexible cyclic forms.

- A three-membered cycle should have no multiple bonds.
- In four-membered rings, only one double bond is allowed.
- Any transition state whose edges (bonds) make up part of an existent ring, can have at most  $\frac{1}{2}$  of its edges in common with this ring. For states of odd length,  $\frac{1}{2}$  means the greatest integer less than  $\frac{1}{2}$ .
- If the cycle involves two rings that are fused, then only one edge from each ring of the fused ring system can contribute to the edges of the proposed cycle; as many single edge contributions as are found are permitted. For example,



is a possible six-membered transition state, but



is disallowed.

- No considerations of steric hindrance are attempted.

*IV.5 Primitive Reactions: Their Motivating Conditions and Resultant Transforms.*

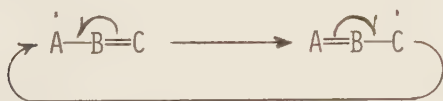
In the forthcoming sections, will be given brief, but complete, descriptions of the primitive reactions of the ION GENERATOR; they will take the form:

- model structure
- qualifying conditions for motivation
- resultant transformation of the model structure
- program names (surveyor, realization, back-reaction), reaction numeral.

These schemata are meant to explain the reactions from a chemical viewpoint. no attempt will be made to indicate how these reactions have been realized in the FORTRAN language. But the program names would enable someone with the program listing of the IG, to discover how these reactions have been implemented. The reaction number is given so that certain options, to be described in Chapter V, can be properly used, e.g., reaction suppressions.

The model structure is the collection of atoms and their interconnections which is sought by the reaction surveyor program. The qualifying conditions are applied by this program also; hence these conditions are explained in terms of the model structure. The resultant transforms are done by the reaction realization routines, inverse transforms by the back-reaction programs.

Anytime that reactions break bonds, they cannot be formed later on again in the same mechanistic sequence. When reactions make double bonds, these can only motivate double bond-triggered reactions. Such controls are required to avoid the following type of situations, illustrated with  $\alpha$ -cleavage.



#### IV.5.1 Ionization of the Molecule

(The many patterns, for which a search can be made, are given in the ionization table of the prologue data set, described in Section V.3.3)

- That atom with the highest position in the ionization table is transformed in one of the following three ways:

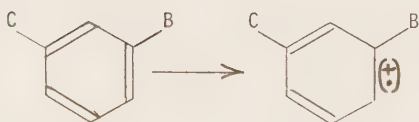
##### Ionization of:



lone pair on atom A.



$\sigma$ -bond between A and B ( $\pi$ -bonds can be ionized also).

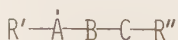


$\pi$ -system, localized near a non-hydrogen atom bond.



PDION,PUION,INVION, 0.

#### IV.5.2 Simple Homolysis ( $\alpha$ -cleavage ).

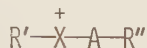


- The radical site on atom A can be deactivated.
- It cannot have been created in the latest step by a hydrogen rearrangement.
- It can be part of a one-electron bond.
- Atoms B and C must have intact outer shells.
- If atom C is an H atom, it cannot belong to a methyl group; it will be cleaved only if the cleavage of all the other  $\alpha$ -bonds will not result in expulsion of a neutral species; the substrate must be the molecular ion, be it modified or not.

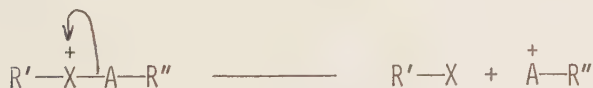
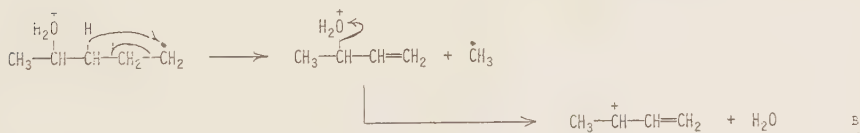
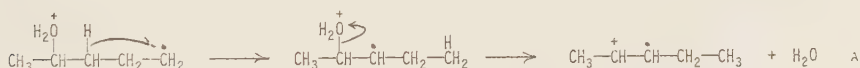


HOMOPR, DOHOMO, INVHOM, 1.

IV.5.3 Heteroatom Heterolysis

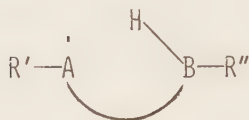


- Atom X, bearing the charge, must be a heteroatom.
- If the substrate structure is the molecular ion, a radical site may also reside on X; otherwise, the heteroatom must be overbonded with respect to its valence.
- Atom A must not have borne the charge in a prior step of the mechanism.
- Atom A must not participate in a multiple bond.
- Expulsion of a neutral fragment through a heterolytic cleavage can only happen once to a given *m/e* value, i.e., no non-fragmenting alteration can precede a heterolytic cleavage of this sort, once the conditions for it to occur have been met. For example reaction A is not allowed, but reaction B is.



HETRPR, DOHETR, INVHET, 2.

IV.5.4 Simple Hydrogen Rearrangement (Cyclic Transition State)

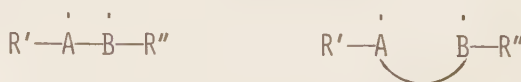


- Only two hydrogen rearrangements are allowed in any one mechanism (optional).
- The last rearrangement must result in a stabilized or deactivated radical.
- Only active radicals can cause rearrangements of this type.
- If the radical is not situated on an electron-pair-bearing atom, or the molecular ion is not the substrate, only weakly-held hydrogen atoms can move, no matter what the size is of the transition state.
- Any weakly held hydrogen atom is allowed to undergo rearrangement, through transition states larger than six membered, if needed.
- No matter which enabling condition is found, the transition state needed for the rearrangement must be flexible.
- If atom A is nitrogen, only weakly-held hydrogen atoms can move.
- Atom B cannot be a methyl carbon.



TRNSPR, DOTRAN, INVTRN, 3.

IV.5.5 Formation of a Bond From Two Radicals

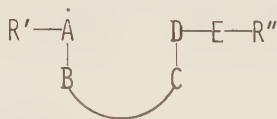


- If a ring will be formed, it must be less than or equal to six membered and must be flexible.



MKBPRD, MKBACT, MKBINV, 4.

#### IV.5.6 Radical Site Motivated Ring Closure



- The number of rings allowed to form is a user option; it is kept usually at one.
- If the radical site at A is stabilized and the substrate is no longer the molecular ion, ring closure will not be performed.
- If the radical site is stabilized in the molecular ion, and the ring size is greater than six, the ring will not be formed.
- Atom D should have an intact outer shell.
- If bonds A-B and C-D are both double, and the substrate is no longer the molecular ion, or if both are double bonds but the required ring is too large, ring closure will not occur.
- Bond D-E must be a single bond.
- Atom E must be outside of the future ring and must have an intact outer shell; it must not be part of a double bond.
- The ring that will form must be flexible.
- To overcome vinylic cleavage when bond D-C is double, the ring must be able to form a resonance hybrid.
- If the ring ends up to be larger than six membered, then bond D-E must be weak.



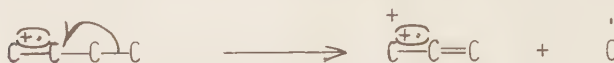
RCPRED, RCACTN, RCINVR, 6.

#### IV.5.7 One-Electron Bond Dismantling



- After having been used as a motivation for a reaction, a one-electron bond must be dismantled. The using reaction will have placed a positive charge in the proper place if it was a radical motivated primitive and a radical site

if it was a charge motivated mechanism. For example, in the following reaction,



homolysis uses the one-electron bond for motivation. The dismantling reaction just removes the one-electron bond to give,



with the bond counts and charges in the correct places. The advantage of this method of bond accounting is that the atom where the charge should be considered to be is fixed by the reaction that is motivated by the one-electron bond, leaving the dismantling chores to this reaction.

- If a single bond has been ionized in the molecular ion, this reaction dismantles the one-electron bond in both of the senses. For example,



- The only bonding requirement is that the bond A-B be fissionable at this point in the mechanism.



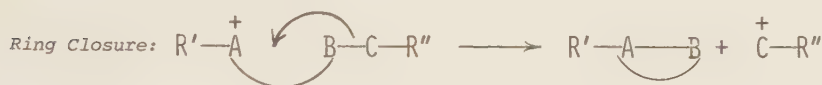
ONEEVP, ONEEVA, ONEEVI, 7.

#### IV.5.8 Non-Heteroatom Heterolysis and Skeletal Rearrangement

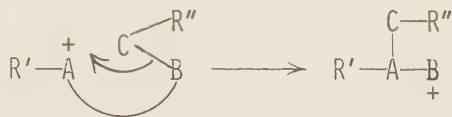


- A maximum of two such reactions can take place in one mechanism; this can be changed by an option.
- No radical sites can be in the substrate, but the motivation can be a one-electron bond.
- Atom A cannot be a heteroatom.
- Atoms B and C cannot be hydrogen atoms.

- When bond B-C breaks it should not be a vinylic cleavage.
- Bond B-C should be a single bond.
- If A is not connected to B, then a ring will form; this ring should be flexible.
- The atom which will end up with the charge must not have had it in a prior step of the mechanism.
- If atom A is connected to atom B then the skeletal rearrangement form cannot occur.

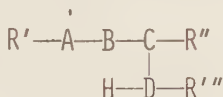


Skeletal Rearrangement:



HTRNGP, HTRNGA, HTRNGI, 8.

#### IV.5.9 Homolysis with Satellite Rearrangement

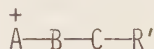


- All of the same conditions for Simple Homolysis, Section IV.5.2, hold as well for this reaction.
- The limit on hydrogen rearrangements has effect on this reaction.
- Atom D cannot be a methyl carbon.
- Atom D should have an intact outer shell and D-H cleavage should not be vinylic in nature.

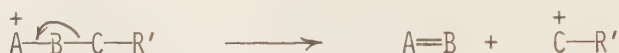


HOMOPR (Second Part), SPHOMA, SPHOMI, 9.

#### IV.5.10 Heterolysis of a Two Atom Functional Group

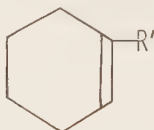


- Atom A must be a heteroatom which is overbonded with respect to its valence.
- Atom B must be the only atom to form a bond with atom A, excluding hydrogen.
- Bond B-C must be a single bond.
- Atom C must never have been charged at a prior time in this mechanism.



COLSSP, COLSSA, COLSSI, 11.

#### IV.5.11 Retro-Diels-Alder

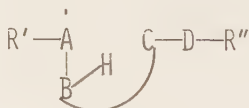


- The double bond should be in a six-membered ring and should not be part of a conjugated double bond system.



RDARPR, RDARAC, RDARIN, 12.

#### IV.5.12 Hydrogen Push Rearrangement



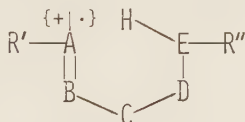
- This rearrangement has the hydrogen atom moving away from the radical, in contrast to the rearrangement of IV.5.4, where the hydrogen atom goes to the radical site.
- The radical on atom A cannot be deactivated.

- Like all rearrangement reactions, the second rearrangement should result in a stabilized radical site.
- Atom B should have a complete outer shell.
- Cleavage of the C-D bond should not be a vinylic cleavage.
- Bond C-D should in all cases be a weak bond; transition state should be flexible.



RCLTHP, RCLTHA, RCLTHI, 13.

IV.5.13 *Double Bond Triggered Hydrogen Rearrangement, Step I (McLafferty Rearrangement)*

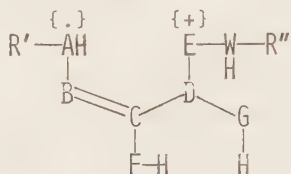


- The A-B double bond triggers this abstraction of a hydrogen atom from atom E.
- If atom A has the positive charge, then the hydrogen can move heterolytically, making E the site of the positive charge. Normally the hydrogen atom is homolytically rearranged.
- Atom E can be a methyl carbon only for a molecular ion substrate.
- Step II is called directly after Step I so that no other primitives can transform the half-reacted result of this step.



MLSTPR, MLSTAC, MLSTIN, 14.

IV.5.14 Double Bond Triggered Hydrogen Rearrangement, Step II (McLafferty Rearrangement)

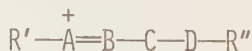


- In all cases, bond C-D will break.
- If a radical site is available on atom A, the various satellite rearrangements possible from W to C, from G to C and from F to D will occur if one more rearrangement is permitted within the two rearrangements limit.
- When the rearrangement is from W to C, then atom C must be a heteroatom.
- Three examples of the results of step II will be shown below, of the many that are possible.



MLS2PR, MLS2AC, MLS2IN, 15.

IV.5.15 Homolytic Overbond Relief of a Heteroatom



- Atom A, bearing the charge, must be overbonded with respect to its valence.
- Atom A and atom B must have a double bond between them.
- Bond C-D must be a single bond.
- When cleavage of C-D occurs it should not be vinylic.
- Atoms C and D must have intact outer shells.



ROVBPR, ROVBAC, ROVBIN, 16.

IV.5.16 *Construct a One-Electron Bond from Adjacent Charge and Radical Sites*  
(Stop Action)

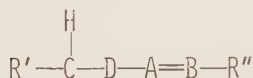


- The prior reaction step cannot be a simple hydrogen rearrangement.
- Atom A should not be an overbonded heteroatom.
- The resulting one-electron bond cannot motivate further reactions.



MOEBPR, MOEBAC, MOEBIN, 17.

IV.5.17 *Bond Replacement by a Hydrogen Atom Bond to an Overbonded Heteroatom*



- The hydrogen rearrangement limit applies to this reaction.
- Atom A must be overbonded having one double bond.
- Atom D must have no double bond so that bond A-D will cleave.
- Atom C must have no double bonds so that the H atom will not make a vinylic bond scission to rearrange.



## Chapter V.

### HOW TO USE THE ION GENERATOR PROGRAM

This chapter of the thesis is meant to serve as a user's guide to the complete ION GENERATOR program. It is intended to be understandable by a mass spectrometrists with a knowledge of FORTRAN. It covers both deck preparation and the interpretation of the results. Information necessary for the implementation of the ION GENERATOR on another machine is to be found in Appendix F.

#### V.1 General Guidelines

The usage of the ION GENERATOR (IG) program is simple, except for the first time, because then the first part of a two-part data deck must be prepared; it is called the prologue data set. The second part of the deck is the structure-spectra sets. The prologue data set rarely changes from run to run of the IG. In a single run, the prologue part is read but once, remaining valuable for all of the given structures. Appendices A, B and C give detailed explanations of the second part of the data deck; here we will concentrate on the prologue, reviewing, as well, structural input and output.

#### V.2 CDC 3800 Control Cards

The IG is kept as an overlay tape, number CCI-2658. After the proper control cards are supplied, the prologue data set is given, followed by the structure-spectra sets. The last structure is followed by the EOF card for the CDC 3800.

The control cards consist of a JOB card, two EQUIP cards and one LOADMAIN card. The required cards are given in detail below. The various parameters given on these cards are described in the user's guide of the Centre Cantonal d'Informatique /60/.

Card Col. :        1                    10                    20

JOB card :    ~~J~~OB,UG012502,DELFINO,5

EQUIP card:    ~~E~~EQUIP,3=RA

EQUIP card: ~~9~~EQUIP,33=(2658SPMAUG0125)  
LOADMAIN card: ~~9~~LOADMAIN,33,5,10000

(~~9~~ is the multiple punch 7 followed by 9)

### V.3 Prologue Data Set

The prologue data set can be divided into three parts: the common set-up options, the periodic table of elements and the ionization-charge localization table.

#### V.3.1 Common Set-Up Options

The first two cards of this subset are variable format cards. The first is used whenever the nodal push-down stacks are printed /Appendix D/; the second card, also a variable format, is used for Diary stack printing. Those given below have been found to be useful.

Nodal Stack : (\* P D S \*,5(I7,3H \* ),\* PRIM\*,I3,\* LVL\*,I5)  
Diary Stack : (\* DIARY\*,I3,4I7,2I4,3I2,4I6,10I2,6I3)

The card that follows contains the many options which govern certain limits on the theory and the amount of intermediate printing to name two categories. A sample options card is given below, 40 columns per line.

col 1: 1CYCLIC 00000 09000 40 5002 11  
col 41: 1 1 00060 19060.5 9999060 1 02

There are twenty-seven options on this card; they will be explained below. The format of the option is given in parentheses.

Column 1 ( L1 ): This option controls the intense trace feature built into the IG. If the user would like to follow each step that the program does on a given structure, a 1 is punched.

Col. 2- 9 ( A8 ): An eight character tag may be given which is put into the disk file records created in Overlay II.

- Cols. 10-12( 13 ): Three digit numbers, starting with the value given here, will be added to the eight digit tag for each of the structures in the run, so that each structure will have a unique tag.
- Column 13( L1 ): If high-resolution masses should be calculated a 1 is given.
- Column 14( L1 ): If metastable masses should be calculated a 1 is given.
- Cols. 15-18( I4 ): The lower bound in ionization probability to be used in a search of the table. A value of 0 is normally used.
- Cols. 19-22( I4 ): The upper bound in ionization probability to be used in a search of the table. A value of 9999 is normally used.
- Cols. 23-26( I4 ): This option, LOWCUT, is the lowest mass ion which should have its provoking mechanism printed in the final output. Any ions coming up with masses inferior to this value are simply discarded. This causes the discrepancy in ion numbering, such that less ions are printed in Overlay III than were numbered because the ones less than LOWCUT were eliminated. A value of 15 would permit every ion to print out.
- Cols. 27-30( I4 ): When a comparison spectrum is supplied, the spectrum can be used to set a threshold mass, below which an ion will not be submitted to further decompositions. The first ion from the spectrum, having a mass greater than LOWCUT, that has a relative abundance (base peak=1000) greater than the value given here, will be the threshold mass if it is greater than the percentage method given below.
- Cols. 31-32( I2 ): The largest number of hydrogen rearrangements which will be allowed in any mechanism. A value of 2 has been sufficient.
- Column 35( L1 ): A stack system trace will be printed if this column has a 1.
- Cols. 36-39( 6X ): Blanks.
- Column 42( L1 ): A one in this column permits no hydrogens to be lost by homolysis, after a rearrangement step. This is always 1.
- Column 43( 1X ): Blank.
- Cols. 44-45( I2 ): The maximum number of ring closures to be permitted in any one mechanism. A value of 1 is sufficient.
- Cols. 46-48( 3X ): Blanks.
- Cols. 49-53( I5 ): The node number, PDEQUIV number, where the intensive traces of the stacks or the step-by-step trace should begin. A value of zero means the beginning. This number does not control the common block dump at every node; see Cols. 68-70.

- Cols. 54-56( I3 ): Maximum length of mechanism. Any value greater than twenty will be ignored, because the program has a physical limitation of 20 steps for any one mechanism; any value 20 or less will have an effect.
- Cols. 57-58( I2 ): The maximum ring size to be permitted for a ring closure or a transition state that does not involve weak bonds. This value is normally left at six.
- Cols. 59-63(F5.1): This is a multiplier value, kept between 0 and 1. The product of this value and the molecular weight will be used for a threshold mass, if the product is greater than the threshold found from the spectral method (See cols. 27-30). We have found 0.5 to work quite well.
- Cols. 64-65( I2 ): The maximum number of fragmentations that can occur in a single mechanistic sequence. Any value of 20 or greater does not effect any control on the program.
- Cols. 66-67( I2 ): The maximum number of steps to be permitted between fragmentations. A value of 20 means no control.
- Cols. 68-70( I3 ): This is the node number after which the intensive common dump should occur. A value of zero means no such dumps should be done, except the few that go along with the other intensive traces.
- Cols. 71-72( 2X ): Blanks.
- Column 73( L1 ): A 1 is punched here in order to have every step of every mechanism printed in the greatest detail possible in Overlay III.
- Cols. 74-75( I2 ): This is the maximum number of carbon skeleton rearrangements which are permitted in any one mechanistic sequence. A value of 2 has worked well.

The next card in order is the reaction suppression card. A one punched in that column of this card, having the same number as that given in chapter IV, will suppress that reaction. A one in column 2 will suppress heteroatom heterolysis for example. The program is normally executed with a blank card here. This card does give the mass spectrometrists the ability to perform experiments showing the effects on the resulting mechanisms and masses generated by leaving out various mechanistic steps.

### V.3.2 Periodic Table Input

As complete a periodic table need be given as will be necessary to process the given structures. But giving a bigger table does not reduce the efficiency, so that a table can be constructed once and for all. The elements are given in the order that they should appear in ionic formulae; i.e., carbon, followed by hydrogen, followed by those atoms for which one has preference. The example set has the order: C,H,D,N,O,Si,P,S,Cl,Br and I.

The first card of the periodic table set is that card describing the atoms that can enter into the same reactions as does the hydrogen atom. The identification letter of these atoms is listed on the card, starting in column 1. If one were going to give a deuterated species to the IG, then one would punch "D" in column 1, leaving the rest of the columns blank. If no such labelled compounds will be given to the IG, then this card should be blank; the IG executes faster if there are no label atoms possible.

The next set of cards, followed by an all blank card to indicate the end of the set, is that set containing the elements and their naturally occurring isotopes, their valence and an indication of it if they have a lone pair of electrons. We have found the following set to be effective.

| Col. 1 | 2     | 3 | 4 | 5 |
|--------|-------|---|---|---|
| C      | 4     | 0 | 0 | 0 |
| H      | 1     | 0 | 0 | 0 |
| D      | 1     | 0 | 0 | 0 |
| N      | 3     | 1 | 0 | 0 |
| O      | 2     | 1 | 0 | 0 |
| Si     | 4     | 0 | 0 | 0 |
| P      | 3     | 1 | 0 | 0 |
| S      | 2     | 1 | 0 | 0 |
| Cl     | 1     | 0 | 0 | 0 |
| Br     | 1     | 0 | 0 | 0 |
| I      | 1     | 0 | 0 | 0 |
| {      | BLANK | } |   |   |

In the first column of a periodic table card, the identifying label is given. Hence the first card has a "C" for carbon. Column 2 is the valence. Column 3 should have a 1 for all the atoms that have lone pair electrons. Column

4 contains the number of isotopes described on the card. For the carbon card there is a 2, which permits the input of  $^{12}\text{C}$  and  $^{13}\text{C}$ . The first isotope to be described should be the most abundant. In the future, the isotopic ion patterns could be calculated by the IG. Columns 15-17 describe the first isotope group of low-resolution mass (3 columns), high resolution mass defect in ten-thousandths of a mass unit (6 columns), and natural abundance in per-thousandths (4 columns). Every 13 columns thereafter are used for succeeding isotope groups. On the carbon card, we find  $^{12}\text{C}$  has a mass defect of zero and natural abundance of 98.9%, with  $^{13}\text{C}$  having a mass defect of +0.0034 and abundance 1.1%. The rest of the table should be self-explanatory.

### V.3.3 Ionization Table Input

The next set of cards in the input stream of the IG is the table of ionizations, given in the decreasing order of probability that the ionization would occur. Below is the complete set that is normally given to the IG.

| <u>Card Number</u> | <u>C a r d C o l.</u> |           |
|--------------------|-----------------------|-----------|
|                    | <u>1</u>              | <u>10</u> |
| 1                  | 16                    |           |
| 2                  | C11600001             |           |
| 3                  | N31500010             |           |
| 4                  | N21400010             |           |
| 5                  | N11300010             |           |
| 6                  | S21200010             |           |
| 7                  | O21100010             |           |
| 8                  | S12040010             |           |
| 9                  | O12030010             |           |
| 10                 | S11020010             |           |
| 11                 | O11010010             |           |
| 12                 | C31001500             |           |
| 13                 | C21001000             |           |
| 14                 | C14000400             |           |
| 15                 | C13000300             |           |
| 16                 | C12000200             |           |
| 17                 | C11000100             |           |

This ionization table contains entries only for C, N, O and S. If compounds containing other atoms were to be run, this table would have to be changed or added to. There is space in the program for 25 ionizations; 16 are given here.

Column 1 contains an atomic label that corresponds to one of those from the periodic table input. Column two is a bond order that will be used to define a bond environment. Column three is a number of the kind of bonds described in column 2. Columns 4 through 7 make up the probability number that the user has assigned to this atom. The scale is relative, not representing any accepted system of ionization probability. Column 8 is the indication that the ionization should occur in the atom's lone pair. Column 9 indicates that ionization of a ring, having the resonance property, should occur.

The IG uses this information in a straightforward fashion. Starting with the most probable type of ionization, the program searches the structure for the bonding environment described on the card. If card numbered six were used, the program would search the structure of the given molecule for a thioether functional group, i.e., a sulfur atom with 2 single bonds, ionizing the lone pair of electrons.



If a bond ionization is demanded by the environment, such as that described on card 13, then the bond order that defined the environment will be ionized to a one-electron bond, in this case a double bond attached to a carbon.



The ionization of a resonance ring, such as a phenyl group, occurs at all of the double bonds, involving carbon atoms to which is attached something other than a hydrogen atom, resulting in a one-electron bond.



The search of sites of ionization is restrained, using the power of ten which the probability value would demand, if it were expressed as a logarithm.

All of the structural environments found in the structure, whose ionization probability is 3 or greater in logarithms to the base ten, are done no matter how many are found. Those with a logarithm less than three are done only if no environments with greater probability were found; if they are required, all of them that are found will be done. The same conditions apply to those with logarithms between 1 and less than 2. Those at the end of the list will be done as a last resort. That type which is first found will be the only environment done, and only three of those. For example, if a molecule contains one tertiary carbon and one secondary carbon, only the bonds of the tertiary carbon will be ionized. If the molecule contains more than three secondary carbons, the first three encountered will be ionized, and then the ionization process would stop.

#### *V.4 Structure-Spectra Data Set(s).*

The second part of the data deck for the IG, is a group of structure-spectra data sets. There can be as many as desired, as long as there remains time from the amount demanded on the job card, and as long as there are enough lines remaining from the amount demanded on the "loadmain" card. These sets each have three parts: an identification or title card, a comparison spectrum and a structure. Because of the nature of the last two, these sets can be of varying length, from a 5 card minimum, up to as many as are necessary.

##### *V.4.1 Identification Card*

The identification card gives an 80 character title to the data set comprising the spectrum and the structure. This title is printed just after the mass-only mass spectrum and just before the mass spectrum. The IG takes no input information concerning the molecule from this card.

##### *V.4.2 The Comparison Spectrum*

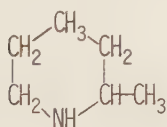
If the actual mass spectrum is available, it can be given to the IG, so that it can print the relative abundance from this given spectrum of the low-resolution mass corresponding to each of the ionic structures predicted. The actual mass spectrum is given in a condensed format. In columns 5 through 8 of the first card, the number of ions to be given is punched in I4 format. The number of ions must be less than or equal to 397, the size of the table reserved

for the spectrum in the program. Directly thereafter on the same card starting in column 9, the ions are given as mass, followed by intensity, both in I4 format. As many eight column sets, as are required by the count, would be used. For a spectrum containing 27 ions, 9 ions would be punched on the first card, 10 on a second, and finally 8 ions on the last card. The relative abundances should be expressed in the "base peak equals 1000" system.

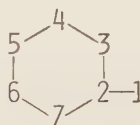
If no spectrum is to be given, a one is punched in column eight of the sole spectrum card. This is taken by the IG to mean no spectrum. Therefore, an actual spectrum must have at least two ions in order to be sensed. The details concerning the spectral search techniques employed by the IG are described in Appendix D.

#### V.4.3 Structure Input

In order to give a molecule to the IG, the non-hydrogen atoms of the structure are numbered in any convenient sequence. The chosen numbering sequence does not affect the results. Below, we show a molecule, 2-methyl piperidine, which has been numbered in the fashion shown to the right.



is numbered as



It can be considered for the purposes of input to be a straight chain, starting with atom 1, the methyl group, and ending with the nitrogen, atom number 7. In addition, atom number 7 is connected to atom number 2 to make the six-membered ring. The cards illustrated below express these two features.

| <u>Card Number</u> | <u>Card Columns</u>                                                                                                                        |
|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------|
| 1                  | <div style="display: flex; justify-content: space-around; margin-bottom: 5px;"> <span>1</span><span>4</span><span>7 8</span> </div> CCCCCN |
| 2                  | 10010071                                                                                                                                   |
| 3                  | 00020071                                                                                                                                   |
| 4                  | 99999999                                                                                                                                   |

Card one gives to each atom number an element label, which corresponds to those given in the periodic table. Column 7 makes atom 7 a nitrogen atom. Card two expresses the straight chain by its one in column one. Chains are assumed to be numbered in sequence; this chain starts with atom 1 (column 4) and ends with atom 7 (column 7), thus having seven members. The bonds between the atoms of the chain are single bonds indicated by the "1" in column 8. Card three indicates one bond (column one), between atoms 2 and 7, and the bond is a single bond (column 8). The last card, number 4, is used to signal the end of the input deck of this structure.

Since the numbering is arbitrary, many input sets are possible for any molecule. But there are usually only a few convenient ones.

#### V.5 Output of the ION GENERATOR PROGRAM.

As was mentioned in section V.3.1, many intense traces are built into the IG. These might be useful to someone trying to learn how the program works. But these outputs are of no interest to the mass spectroscopist user of the program. They will not be described here. They are for the most part self-explanatory. What will be described will be the mass-only mass spectrum and the detailed printout of each mechanism generated.

##### V.5.1 The Mass-Only Mass Spectrum

The mass-only mass spectrum of cyclohexene, as it was produced by the IG, is given below.

|    |    |    |       |
|----|----|----|-------|
| 1  | 0  | 82 | C6H10 |
| 2  | 4  | 81 | C6H9  |
| 3  | 12 | 81 | C6H9  |
| 4  | 9  | 67 | C5H7  |
| 5  | 17 | 67 | C5H7  |
| 6  | 8  | 55 | C4H7  |
| 7  | 16 | 55 | C4H7  |
| 8  | 6  | 55 | C4H7  |
| 9  | 14 | 55 | C4H7  |
| 10 | 11 | 54 | C4H6  |
| 11 | 19 | 54 | C4H6  |
| 12 | 1  | 54 | C4H6  |
| 13 | 5  | 53 | C4H5  |
| 14 | 13 | 53 | C4H5  |
| 15 | 10 | 53 | C4H5  |
| 16 | 18 | 53 | C4H5  |
| 17 | 2  | 41 | C3H5  |
| 18 | 3  | 41 | C3H5  |

|    |    |         |
|----|----|---------|
| 19 | 7  | 41 C3H5 |
| 20 | 15 | 41 C3H5 |

This output is self-explanatory except for column two. Column one is the number of the mechanism in the output of detail mechanism reports, counting from the end to the beginning. In the case of cyclohexene, mechanism number 20 follows just after the mass-only mass spectrum. Column two is the order that the mechanisms were generated by the first overlay program. Since some mechanisms are ignored, those that generate ions that are lower than  $m/e$  40 in low resolution units, these numbers are not the same as the mechanism number. Column three is the  $m/e$  of the ionic structure that was generated. Finally, column four is the ionic formula.

#### V.5.2 Mechanism Reports

The most important output of the IG program is the mechanism report. The example, given below, is the report of the mechanism corresponding to the classical McLafferty rearrangement of a hydrogen in the molecular ion of dodecane-6-one ( $m/e$  114); it was assigned mechanism number 3, out of 154 ionic structures generated, giving rise to 36 different ionic formulae, and 32 different low-resolution masses. Fifteen structures were not printed having less than  $m/e$  40.

```
MECHANISM 3 MASS 114 ION FORMULA C7H14O INTENSITY(GIVEN) 60
MECHANISM HAS 3 STEPS.
```

```
PLEASE SEE UNDER LRM 184 THE MECHANISM NUMBERED 1.
```

```
A 6 MCLAFFERTY, EQVHYD 0 GOES H(.) FROM 4 TO 13 AND DOUBLE BOND 13-7 ENDS UP
  AT 7-6 and 6-5 BREAKS IN NEXT STEP.
```

```
O(+.)H-C=-CH2-CH2-C(.)H-CH2-CH2-CH3CH2-CH2-CH2-CH2-CH3 C12H24O
ATOM NUMBERS - 13 7 6 5 4 3 2 1 8 9 10 11 12
```

```
MC LAFFERTY HOMOLYSIS, RING 0 EQUIV 0 GOES FROM 0 TO 0 NEW BOND 0 0
  SITE POS AND TYPE 4 1
```

```
IONIC SPECIES . .
O(+.)H-C=-CH2-CH2-CH2-CH2-CH3CH2 C7H14O
ATOM NUMBERS - 13 7 8 9 10 11 12 6
NEUTRAL SPECIES. . .
CH3-CH2-CH2-CH=CH2 C5H10
ATOM NUMBERS - 1 2 3 4 5
IONIC MASS 114 NEUTRAL SPECIES MASS 70
```

THIS MECHANISM, NUMBERED 3, GIVES RISE TO THE IONIC STRUCTURE WHICH IS THE PRECURSOR OF THE IONS IN THE FOLLOWING MECHANISMS. . . 43 6, 44 5, 58 4.

\*\*\*\*\* E N D O F M E C H A N I S M \*\*\*\*\*

Each primitive reaction step of the mechanism is explained with every available detail. In this reaction, we see three steps: ionization, McLafferty step 1, and McLafferty step 2. The fragmentation is signaled by the small group of printed lines containing "IONIC MASS" and "NEUTRAL SPECIES MASS". At each stage of the mechanism, the structure of the ion after the transformation is represented in the linear structure notation explained in Appendices A, B and C, and also in the following section. At the very end of the report, mention is made of all the daughter ions to the current end ion; they are given as "LRM Mechanism", i.e., the low resolution mass group in which they are found, and the mechanism number that they were assigned.

The set of mechanism reports are grouped by low-resolution mass in ascending order. All mechanisms with the same mass are grouped together by ionic formula. Inside these formula groups, the mechanisms are ordered by the number of primitive steps in the sequence. Mechanisms with the same length, with the same reaction steps operating on the same masses, are tagged as equivalent mechanisms.

#### *V.5.3 Linear Notation for Structural Outputs*

The linear notation, used in the mechanism report for representing intermediate structures, is a variant of the DENDRAL notation /A:39/. Certain changes have been made to adopt the notation to this particular problem. Appendices A, B and C have various stages of explanation, covering the amount needed for the understanding of each article. But since those articles have appeared, some additional features have been included; therefore we will give a complete explanation of the current version of the ION GENERATOR's linear notation for structural output.

Consider the simple case of methane, given below in the linear notation.



The H atoms attached to an atom are indicated immediately after the atom; if the number of hydrogen atoms exceeds one, the "H" is followed by their number. The linear notation of iso-propylamine exhibits an additional property.



It is that an atom is followed by all of its efferent bond, followed in turn, but in the same order, by the radicals that are attached to these bonds. When a

radical, attached to an efferent bond, is itself made up of more than one radical, another rule is needed. The linear notation of methyl-iso-propyl-sec-butylamine given below exhibits this rule.

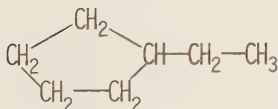
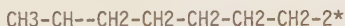
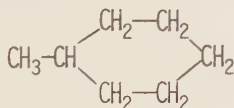


Once a radical is started, it is completely described before going on to the next radical attached to the efferent bonds of the preceeding atom. For acyclic structures, these are all the rules that are needed for neutral, non-radical species. If an atom has either a charge or a radical site or both thereon, it is followed by either (+), (.) or (+.), which is, in turn, followed by the attached hydrogen atoms. When a bonding electron is considered to be ejected, we have a one-electron bond, which is symbolized by  $\Lambda$ . For propene, with a missing  $\pi$ -electron, we would write the following linear formula.

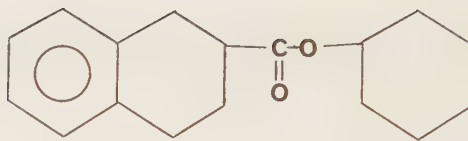


In the brackets is the order of the bond "underneath" the one-electron bond.

For the representation of ring structures, a final rule must be used. When a bond in a radical is connected to an atom that has been written down already, the ordinal number of this atom in the sequence is written just after the bond; i.e., the bond refers back to this atom, forming the ring. At the place in the notation, where the radical should have been described, whose efferent bond was taken up by the back reference, an asterisk is printed. Two examples are given, methyl cyclohexane and ethyl cyclopentane, each of which has one ring.



When more than one ring is found in the structure, the asterisks make a ring count for the structure.



CH=-CH-C=CH<sub>2</sub>-CH--C=O-CH--CH<sub>2</sub>-CH<sub>2</sub>-CH<sub>2</sub>-CH<sub>2</sub>-CH<sub>2</sub>-8\*OCH<sub>2</sub>-CH<sub>2</sub>-C=3-CH=CH-1\*\*

Several representations for a given structure are possible by this method;  
i.e., it is not a canonical representation.

## Chapter VI.

### RESULTS AND DISCUSSION

Appendices A through E of this thesis are scientific articles that have been either published (Appendices A through D) or which will be submitted for publication (Appendix E). The greater part of the results of this work have been discussed in these articles. Some of the most recent results will be discussed here.

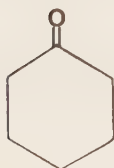
#### *VI.1 The Early Efforts*

When it was decided to write a heuristic program to operate with the electron-bookkeeping theory, it was assumed that the heuristic search of a mechanism space with only simple primitive reaction steps was an ultimately realizable task. This early optimism is reflected in the article given as Appendix A, the preparation of which gave us the opportunity to review the complete field of mass spectral computer programming. This review acted as a natural vehicle for a thorough introduction of the early ideas of the ION GENERATOR, such as, ionization, structural input, structural output, and mechanism construction to name some of the major points. At that time, the hope was to be able to construct all possible combinations of simple mechanistic steps to explain the results of deuterium-labelling experiments.

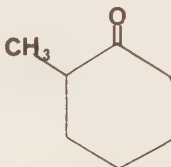
When the article given in Appendix B appeared, the results were quite encouraging. The article is based on a study of the charge localization concept's implementation in the IG. With the structure of 5-diethylamino-pentane-2-one, the IG proposed 2 molecular ions, from which it found 37 primary ions, 22 of them had the charge on the nitrogen atom, and the other 15 had it on the oxygen atom. The program could only, at that time, accept acyclic compounds. The primitives, known for constructing primary ions, numbered three excluding ionization. They were simple homolysis, ring transition rearrangements of hydrogen atoms, and bond formation between two adjacent radical sites. There were no primitives based upon heterolytic transformations.

Encouraged by the fact that all of the accepted mechanisms for amines (which were demanded by the structure of 5-diethylaminopentane-2-one)

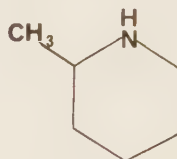
and ketones, which could be constructed from the basis set of primitives, were constructed, the next efforts were made in two directions: to enable the program to operate in the realm of cyclic structures, and to extend the set of primitive reaction steps. Thus the work described in Appendix C shows the IG functioning with the cyclic compounds, shown below.



*Cyclohexanone*



*2-Methyl Cyclohexanone*



*2-Methyl Piperidine*

The mechanisms, shown therein, were constructed with the aid of a fourth additional primitive, homolytic ring closure. At that point, primary ions were still the only one that could be studied. But all of the primary ions of any importance, requiring heterolysis, were to be found in the output.

Because the program contained sufficient structural input and output tools, but did not contain any heterolytic primitive reaction steps, plans were made to study ions to any depth with the then current primitives along with certain heterolytic steps. It was then that it became clear, because of the vast numbers of mechanisms that were generated for simple molecules, that the program as it had evolved was not a good vehicle for the study of reactions constructed from the highly basic primitives. In order to render these early efforts useful for organic mass spectrometry, a shift of emphasis was needed.

## *VI.2 The Change in Strategy*

The new strategy for the program was to develop this collection of functioning programs to do another important task in mass spectrometry, which is closely related, the prediction of the ionic structures of a mass spectrum based on the electron-bookkeeping theory; we would not generate every conceivable mechanism. What resulted, essentially, was the primitive reaction set given in Chapter IV. When the manuscript given in Appendix E was written, the reactions involving one-electron bonds and conjugated double bonds, along with the retro-Diels-Alder reaction were not fully worked out. These features will be discussed in a later section of this chapter. In the manuscript we

see some of the encouraging results that the improved primitives bring.

The differences, between the primitives of the old strategy and the primitives used in the new, can be seen from the following examples of  $\alpha$ -cleavage of the old definition and that of the new.

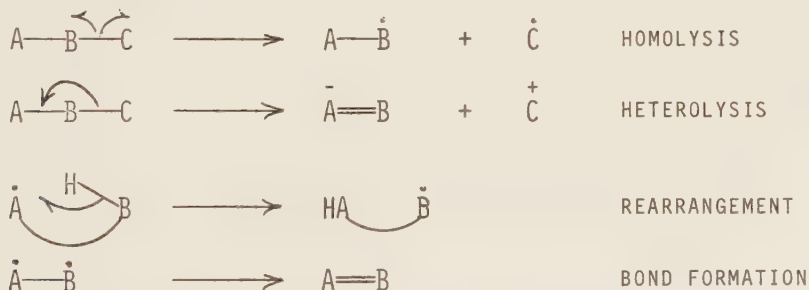
*Old Definition:*



*New Definition:*



The new definition is the old  $\alpha$ -cleavage mechanistic step with the make-a-bond step always done together; i.e., a concerted set of simple operations. One can consider that the whole of electron-bookkeeping theory contains just four indivisible operations:



Essentially, we have defined sixteen concerted sets of these four operations which will, in many cases, generate most of the masses of a mass spectrum, so doing with generally accepted mechanistic steps. Also, few ions that are not in the spectrum are generated.

These sets of concerted operations have the additional property that they may be combined as often as the structural environment dictates, no matter how many neutral species have been expelled. Although not all ions can be generated by this small set of primitive reactions, good results are obtained. The IG, in each of the case studies shown in Appendix E, has operated on daughter ions of the molecular ions, and even on second generation ions. In fact, the

only depth limit which can be changed at any time by the user, is that the ion on which the IG works must be over one half of the molecular weight. The number ( $N$ ) of different mechanisms that result for each compound and the percent (%) of the ion current covered by the resulting masses are shown below.

|                              | $N$ | %  |
|------------------------------|-----|----|
| Methyl-n-propyl-n-butylamine | 14  | 80 |
| Ethyl sec-Butyl Ether        | 15  | 85 |
| Octane-3-ol                  | 83  | 91 |
| Dodecane-6-one               | 139 | 88 |

Each of these structures caused the IG program to propose those accepted mechanisms, for each of the ions, where studies have lead to them. For others, chemically acceptable mechanisms are usually proposed.

#### VI.3 Recent Studies With the ION GENERATOR Working With Cyclic Compounds.

The recently developed features of the IG, which haven't yet been discussed in a published article are:

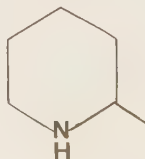
- The motivation of reactions by one-electron bonds.
- The use of primitives which operate solely on cyclic structures.
- The effect of conjugated double bond systems on double bond triggered reactions.
- The ionization of rings with resonance, such as benzene.

Five different cyclic compounds have been successfully fragmented by the IG, using one or all of the above features. These compounds are:

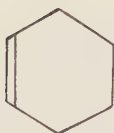
Benzene



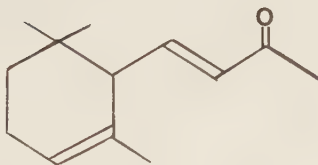
2-Methylpiperidine



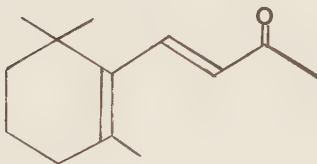
Cyclohexene



$\alpha$ -Ionone

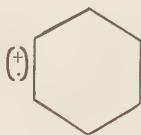


$\beta$ -Ionone

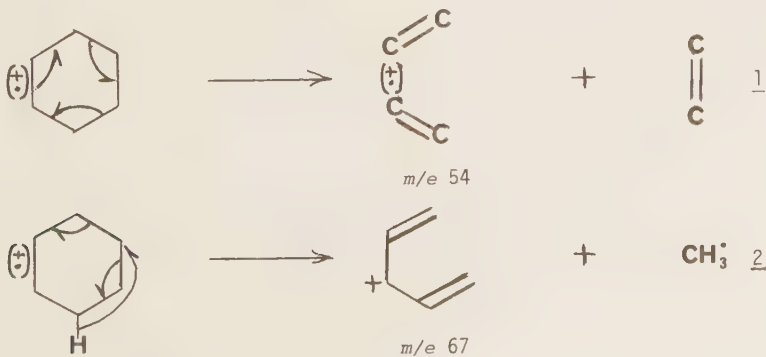


### VI.3.1 Reactions Motivated by One-Electron Bonds

The only way that cyclohexene can be ionized by the IG is by the removal of a  $\pi$ -electron, giving rise to a structure containing a one-electron bond.



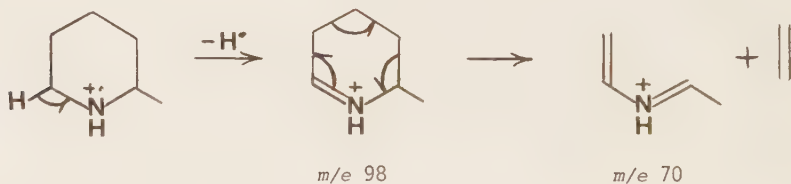
From the output of cyclohexene, we find the following reactions which use the ionized double bond as a driving force



In mechanism 1, the ionized property of the double bond was not important. It is only that the double bond exists in a six-membered ring which motivates the retro-Diels-Alder (RDA) reaction. The  $\{M - 15\}^+$  ion in cyclohexene is the base peak; it is interesting to note that the RDA (mechanism 1) and mechanism 2, which give rise to the two strongest ions in the spectrum, each result from three homolyses, according to the IG; the two mechanisms have no homolytic cleavages in common.

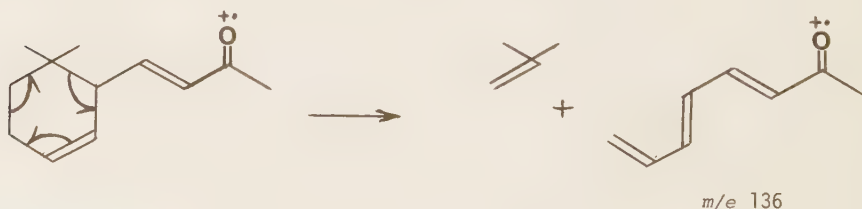
### VI.3.2 Reactions Involving Six-Membered Rings

The RDA reaction is the only primitive in this class of reactions, for the IG. In mechanism 1 of cyclohexene, we saw the reaction occur to a double bond activated by ionization. In 2-methylpiperidine an RDA is done by the IG on the  $\{M - 1\}^+$ , as well as on the  $\{M - 15\}^+$  ion.



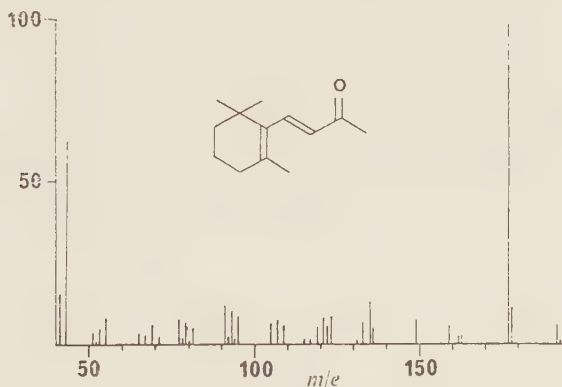
In this mechanism, we have an RDA occurring from a double bond which did not exist in the molecular ion; this double bond, in the neighborhood of the charge localized on the nitrogen atom was formed by the IG.

In  $\alpha$ -ionone, the RDA reaction occurs in the molecular ion, at a distance from the sight of ionization.

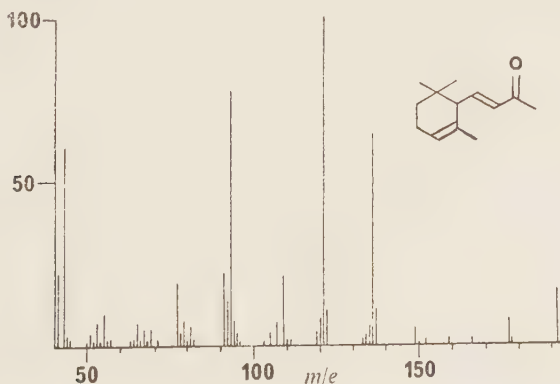


### VI.3.3 The Effect of Double Bond Conjugation

Thomas, Willhålm and Müller /62/ observed the effect of conjugation in doubly unsaturated, carbonyl compounds, two of which were  $\alpha$ -ionone and  $\beta$ -ionone. In  $\beta$ -ionone, shown below with its spectrum (70 eV)/62/ are three conjugated double bonds, one of which is in a six-membered ring.

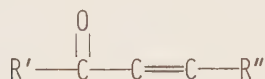


The  $m/e$  164 ion (loss of  $C_2H_4$  through an RDA reaction) is absent in this spectrum, whereas in the spectrum of  $\alpha$ -ionone (70 eV) /62/ shown below, the corresponding loss of  $CH_2=C(CH_3)_2$  leads to  $m/e$  136 with a relative intensity of 60%.



The sole difference between these two molecules is the position of the ring double bond. In one case it is conjugated, in the other case it is not; the relative simplicity of the  $\beta$ -ionone spectrum contrasts with the rather complex  $\alpha$ -ionone spectrum.

Djerassi /52/ discusses a similar effect in  $\alpha,\beta$ -unsaturated ketones, having the general structure given below.



In these compounds, the McLafferty rearrangement is suppressed on the R' side. This is a similar effect to the one pointed out above for the ionones. In the IG, these two effects are taken into account simply by the rule that a conjugated double bond cannot trigger a reaction. The only two double bond triggered primitives are the McLafferty and the retro-Diels-Alder. The masses and frequency of generation of the various masses of  $\alpha$ -ionone and  $\beta$ -ionone are given together in Table VI-1.

TABLE VI-1

Masses and Generation Frequency From the IG.

| <u>m/e</u> | <u>Freq. <math>\alpha</math></u> | <u>Freq. <math>\beta</math></u> | <u>m/e</u> | <u>Freq. <math>\alpha</math></u> | <u>Freq. <math>\beta</math></u> |
|------------|----------------------------------|---------------------------------|------------|----------------------------------|---------------------------------|
| 43         | 3                                | 1                               | 124        | 4                                |                                 |
| 58         | 8                                | 6                               | 125        | 1                                |                                 |
| 83         | 8                                | 4                               | 126        | 1                                |                                 |
| 84         | 1                                |                                 | 136        | 4                                |                                 |
| 93         | 3                                |                                 | 149        | 6                                | 7                               |
| 98         | 8                                | 4                               | 151        | 1                                |                                 |
| 99         | 2                                |                                 | 164        | 4                                | 1                               |
| 109        | 17                               |                                 | 177        | 56                               | 14                              |
| 110        | 1                                |                                 | 192        | 1                                | 1                               |
| 121        | 4                                | 1                               |            |                                  |                                 |
|            |                                  |                                 | TOTAL      | 133                              | 39                              |

The differences between these two results can be attributed mostly to the fact that in  $\beta$ -ionone, the double bonds are fully conjugated.

#### VI.3.4 Ionization of Resonance Rings

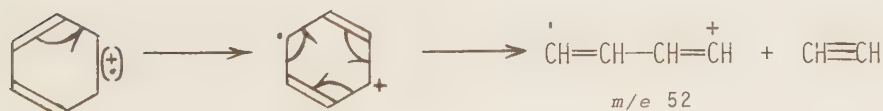
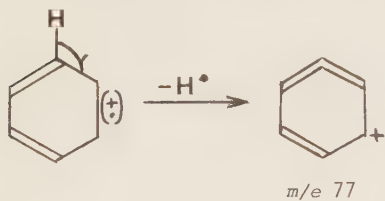
The only structure that has been given to the IG, that has had a resonance ring, has been benzene. The output shows that there is work to be done on the IG in the area of primitive reactions of aromatic systems. Table VI-2 gives the masses and generation frequency for benzene.

TABLE VI - 2.

Masses and Generation Frequency for Benzene

| <i>m/e</i>       | 26 | 52 | 77 | 78 |
|------------------|----|----|----|----|
| <i>Frequency</i> | 6  | 12 | 6  | 3  |

The following three reactions give an example of the generation of each ion; the *m/e* 78 ions arise from the ionization reaction.



#### VI.4 Possibilities for Future Research

Since new primitives can be installed in the IG in a facile manner, studies of the IG's operational characteristics on many compounds could lead to a better understanding of the mass spectrometry of organic molecules. This understanding would be expressed as new primitives made up of combinations of the four basic operations of the electron-bookkeeping theory. Thus the program could become, a step at a time, a better expression of the combined knowledge of mass spectrometry. We have made but a small step in that direction.

Another related area of research is the development, perhaps by similar techniques, of a true mass spectral prediction program. Such a program, given a structure, would manipulate it to find not only the masses that one might expect, but also would give some expression of the relative abundance. It surely would be a difficult task, but such a program would be worth the effort, considering the value that it would have to mass spectrometry.

The field is only just opening for such highly complex programs which do structure manipulations. Now that many of the needed tools have been developed, other workers should find such research rewarding in terms of the insights that are gained when a computer can be taught to do chemical reasoning. The effects of a chemical "notion" on the base of chemical knowledge are instantly known, so that such notions can be refined to high sophistication.

Chapter VII.

B I B L I O G R A P H Y

- /1/ "FORTRAN, 3400, 3600 and 3800 Computer Systems Reference Manual", Pub. No. 60132900, Control Data Corporation, St. Paul, Minnesota 1966.
- /2/ "LISP 1.5 Programmer's Manual", John McCarthy, Paul W. Abrahams, Daniel J. Edwards, Timothy P. Hart, Michael I. Levin, MIT Press, Cambridge 1965.
- /3/ "Computer-Assisted Design of Complex Organic Syntheses", E.J. Corey and W. Todd Wipke, Science 166, 178 (1969).
- /4/ "Computer-Assisted Synthetic Analysis. Facile Man-Machine Communication of Chemical Structure By Interactive Computer Graphics", E.J. Corey, W. Todd Wipke, R.D. Cramer III and W. Jeffrey Howe, J. Amer. Chem. Soc. 94, 421 (1972).
- /5/ "Computer-Assisted Synthetic Analysis for Complex Molecules. Methods and Procedures for Machine Generation of Synthetic Intermediates", E.J. Corey, Richard D. Cramer III and W. Jeffrey Howe, J. Amer. Chem. Soc. 94, 440 (1972).
- /6/ "An Algorithm for Machine Perception of Synthetically Significant Rings in Complex Cyclic Organic Structures", E.J. Corey and George A. Petersson, J. Amer. Chem. Soc. 94, 460 (1972).
- /7/ "Automatic Problem-Solving Applied to Synthetic Chemistry", Malcolm Bersohn, Bull. Chem. Soc. Jap. 45, 1897 (1972).
- /8/ "Applications of Artificial Intelligence for Chemical Inference. Part 8. An Approach to the Computer Interpretation of the High-Resolution Mass Spectra of Complex Molecules. Structure Elucidation of Estrogenic Steroids", D.H. Smith, B.G. Buchanan, R.S. Englemore, A.M. Duffield, A. Yeo, E.A. Feigenbaum, J. Lederberg, C. Djerassi, J. Amer. Chem. Soc. 94, 5962 (1970).

- /9/ "Heuristic Theory Formation: Data Interpretation and Rule Formation", B.G. Buchanan, E.A. Feigenbaum and N.S. Sridharan in: Machine Intelligence 7, Edinburgh University Press (1972) pp. 267-90.
- /10/ "Applications of Artificial Intelligence for Chemical Inference. IX. Analysis of Mixtures Without Prior Separation, as Illustrated for Estrogens", D.H. Smith, B.G. Buchanan, R. S. Englemore, H. Adlercreutz, C. Djerassi, J. Amer. Chem. Soc., in Press.
- /11/ "Applications of Artificial Intelligence for Chemical Inference. XI. Analysis of Carbon-13 Nuclear Magnetic Resonance Data for Structure Elucidation of Acyclic Amines", R.E. Carhart and C. Djerassi, J. Chem. Soc. (Perkin II) 13, 1753 (1973).
- /12/ "Applications of Artificial Intelligence for Chemical Inference. X. INTSUM. A Data Interpretation and Summary Program Applied to the Collected Mass Spectra of Estrogenic Steroids", D.H. Smith, B.G. Buchanan, W.C. White, E.A. Feigenbaum, J. Lederberg and C. Djerassi, Tetrahedron, in Press.
- /13/ "Applications of 'Artificial Intelligence' for Chemical Inference. VI. Approach to a General Method of Interpreting Low-Resolution Mass Spectra with a Computer", A. Buchs, A.B. Delfino, A.M. Duffield, C. Djerassi, B.G. Buchanan, E.A. Feigenbaum and J. Lederberg, Helv. Chim. Acta 53, 1394 (1970).
- /14/ "Computer-Aided Interpretation of Mass Spectra. Part 3. A Self-Training Interpretive and Retrieval System", K-S. Kwok, R. Venkataraghavan, F.W. McLafferty, J. Amer. Chem. Soc. 95, 4185 (1973).
- /15/ "A Self-Training Interpretive and Retrieval System for Unknown Mass Spectra", F.W. McLafferty, R. Venkataraghavan, K-S. Kwok, G. Pesyna, 6th Int. Mass Spectrometry Conf., Edinburgh 10-14 Sept. 1973, Abs. No. 110, 79 (1973).
- /16/ "Computer Analysis of Mass Spectral Information", K.S. Kwok, R. Venkataraghavan, F.W. McLafferty, 19th Annual Conf. on Mass Spect. and Allied

Topics. Atlanta, Georgia, USA 2-7 May 1971, Paper D6, 70 (1971).

- /17/ "Spectroscopic Data Analysis by Computerized Pattern Recognition Techniques", P.C. Jurs, Procs. 11th Nat. Mtg., Soc. for Appl. Spectrosc., Dallas, Texas, USA, 10-15 Sept. 1972, Abst, No. 44, 45 (1972)
- /18/ "Multicategory Prediction Using Arrays of Binary Pattern Classifiers", W.L. Felty, P.C. Jurs, Anal. Chem. 45, 885 (1973).
- /19/ "Application of a Complex-Valued Nonlinear Discriminant Function to Low-Resolution Mass Spectra", J.B. Justice, D.N. Anderson, T.L. Isenhour, J.C. Marschall, Anal. Chem. 44, 2087 (1972).
- /20/ "Information Content of Mass Spectra as Determined by Pattern Recognition Methods", J.B. Justice and T.L. Isenhour, Anal. Chem. 46, 223 (1974).
- /21/ "Artificial Intelligence", J.T. Clerc, P. Naegeli, and J. Seibl, Chimia 27, 639 (1973)
- /22/ "Alternate Representation of Mass Spectra for the Spectral Identification Problem. (Pattern Recognition)", C.F. Bender, H.D. Shepherd and B.R. Kowalski, Anal. Chem. 45, 617 (1973).
- /23/ "Construction of Optimum Variables for Spectral Interpretation (Pattern Recognition)", C.F. Bender and B.R. Kowalski, Anal. Chem. 45, 590 (1973).
- /24/ "Hadamard Transform and Spectral Analysis by Pattern Recognition", B.R. Kowalski and C.F. Bender, Anal. Chem. 45, 2234 (1973).
- /25/ "Automatic Classification of Mass Spectra by Means of Digital Learning Nets", T.J. Stonham, I. Aleksander, M. Camp, M.A. Shaw and W.T. Pike, Electron Lett. 9, 391 (1973).
- /26/ "Applications of Artificial Intelligence to the Interpretation of High Resolution Mass Spectra", D.H. Smith, A.M. Duffield and C. Djerassi, 20th Annual Conf. on Mass Spect. and Allied Topics, Dallas, Texas USA, 4-9 June 1972 Amer. Soc. for Mass Spect. and ASTM Committee E-14, Paper R8, 372 (1972).

- /27/ "Computer-Aided Interpretation of Spectra for the Structure Elucidation of Organic Compounds", J.T. Clerc, F. Erni, C. Jost, J. Meili, P. Nageli and R. Schwarzenbach, Z. Anal. Chem. 264, 192 (1973).
- /28/ "Processing of Mass Spectra for Abridgment of Excess Information", B.N. Kononov, Yadern. Priborostr. 15, 75 (1971).
- /29/ "The Heuristic DENDRAL Program for Explaining Empirical Data", B.G. Buchanan and J. Lederberg, in Information Processing 71, Ed. C.V. Freiman, North-Holland Publishing Co., Amsterdam, 179 (1972).
- /30/ "Automated Chemical Structure Analysis of Organic Compounds. A Computer Program for Analysis of Small-Sized Compounds", S. Sasaki, Y. Kudo, S. Ochiai and I. Fujioka, Bunseki Kagaku (Japan Analyst) 22, 25 (1973).
- /31/ "Computer Programs that Generate Hypotheses in Organic Chemistry", B.G. Buchanan, R.S. Englemore and E.A. Feigenbaum, Trans. Am. Nucl. Soc. 15, 778 (1972).
- /32/ "Structure Determination of Organic Molecules by Parametric Learning Methods", J. Franzen and H. Hillig, 6th Int. Mass Spectrometry Conf., Edinburgh 10-14 Sept. 1973, Abst. No. 109, 78 (1973).
- /33/ "Computer Interpretation of Mass Spectra", D. Koo and R.D. Sedgwick, 6th Int. Mass Spectrometry Conf., Edinburgh 10-14 Sept. 1973, Abstr. No. 115, 81, (1973).
- /34/ "An Application of Artificial Intelligence to Mass Spectrometry", B.G. Buchanan, Amer. Chem. Soc., 162nd Nat. Mtg., Washington D.C. USA, 12-17 Sept. 1971, Abstr. N. Chlt. 19, (1971).
- /35/ "Possibilities of a Mass Spectrometer-Computer System for Structure Analysis of Organic Molecules", H.A. van't Klooster, Chem. Weekbl. 67, 9 (1971).

- /36/ "Further Aspects of the Naive Analysis of Structure", D.H. Robertson and R.I. Reed, 19th Annual Conf. on Mass Spect. and Allied Topics, Atlanta, Georgia USA, 2-7 May 1971, Amer. Soc. for Mass Spect. and ASTM Committee E-14, Paper D5, 68 (1971).
- /37/ "Computer Techniques for Identifying Low-Resolution Mass Spectra", S. Grotch, 19th Annual Conf. on Mass Spect. and Allied Topics, Atlanta, Georgia USA, 2-7 May 1971, Amer. Soc. for Mass Spect. and ASTM Committee E-14, Paper D7, 72 (1971).
- /38/ "Computerized Analysis of Spectroscopic and Chemical Data by Pattern Classification Techniques", L.E. Wangen, Diss. Abstr. Order No. 72-7428 32, 4460B (1971).
- /39/ "The Heuristic DENDRAL Program for Explaining Empirical Data", B.G. Buchanan, J. Lederberg and E.A. Feigenbaum, AD-730506 CS-203 AIM-141 Dept. of Computer Sci., Stanford Univ., Calif. USA (1971).
- /40/ "The Application of Artificial Intelligence in the Interpretation of Low-Resolution Mass Spectra", A. Buchs, A.B. Delfino, C. Djerassi, A.M. Duffield, B.G. Buchanan, E.A. Feigenbaum, J. Lederberg, G. Schroll, G.L. Sutherland, in Advances in Mass Spectrometry. Procs. Conf. Brussels 31 Aug.-4 Sept. 1970, Inst. Petroleum London and Elsevier Amsterdam and New York 5, 314 (1971)
- /41/ "An Application of Artificial Intelligence to the Interpretation of Mass Spectra", B.G. Buchanan, A.M. Duffield, A.V. Robertson, in Mass Spectrometry. Techniques and Applications, Ed. G.W.A. Milne, Wiley-Interscience, New York, 121 (1971).
- /42/ "An Attempt to Apply Computerized Pattern Recognition Systems to Analyze Mass Spectra", S. Sasaki and H. Abe, Mass Spect. Japan 20, 131 (1972).
- /43/ "Preliminary Classification of Mass Spectral Patterns Using a Simplified Learning Machine", R.J. Mathews, Aust. J. Chem. 26, 1955 (1973).

- /44/ "Interpretation of Mass Spectra with Adaptive Linear Classifiers", K. Varmuza, P. Krenmayr, Fresenius' Z. Anal. Chem. 226, 274 (1973).
- /45/ "Identification and Qualitative Analysis of Organic Compounds by Low Resolution Mass Spectra Using the Method of Successive Pattern Identification", M.S. Khots, Zh. Anal. Khim. 28, 797 (1973).
- /46/ Doctoral Dissertation, H.M. Rosenstockm University of Utah, Salt Lake City, Utah, (1952).
- /47/ "Introduction to Mass Spectrometry and Its Applications", R. W. Kiser, Prentice-Hall, Englewood Cliffs, New Jersey, (1965).
- /48/ "Improved Discriminant Training and Feature Extraction for the Generation of Simulated Mass Spectra of Small, Organic Molecules", J. Schechter and P.C. Jurs, Appl. Spectrosc. 27, 225 (1973).
- /49/ "Generation of Simulated Mass Spectra of Small Organic Molecules by Computerized Pattern Recognition Techniques", J. Schechter and P.C. Jurs, Appl. Spectrosc. 27, 30 (1973)
- /50/ "Interpretation of Mass Spectra of Organic Compounds", H. Budzikiewicz, C. Djerassi and D.H. Williams, Holden-Day, San Francisco (1964).
- /51/ "Structure Elucidation of Natural Products by Mass Spectrometry, Vols. I and II", H. Budzikiewicz, C. Djerassi and D.H. Williams, Holden-Day, San Francisco (1964).
- /52/ "Mass Spectrometry of Organic Compounds", H. Budzikiewicz, C. Djerassi and D.H. Williams, Holden-Day, San Francisco (1967).
- /53/ "Artificial Intelligence, The Heuristic Programming Approach", J. R. Slagle, McGraw-Hill, New York (1971).
- /54/ "Molecular Orbital Approach to the Interpretation of Organic Mass Spectra. Theoretical Treatment of the Fragmentation Rules in the Mass Spectrometry of Organic Compounds", H. Ichikawa and M. Ogata, Bull. Chem. Soc. Jap. 46, 1873 (1973)

- /55/ "Mass Spectrometric Analysis. Molecular Rearrangements", F.W. McLafferty, Anal. Chem. 31, 82 (1959).
- /56/ "Introduction to Mass Spectrometry. Instrumentation and Techniques", John Roboz, John-Wiley (Interscience) New York (1968).
- /57/ "Interpretation of Mass Spectra. An Introduction", F.W. McLafferty, W.A. Benjamin, Inc., New York, Amsterdam (1966).
- /58/ "Mass Spectrometry of Organic Ions", Ed. F.W. McLafferty, Academic Press, New York, London (1963).
- /59/ "Mass Spectrometry of Heterocyclic Compounds", Q.N. Porter and J. Baldas, Wiley-Interscience, New York, London, Sydney, Toronto (1971).
- /60/ "Manuel de l'Utilisateur", Centre Cantonal d'Informatique, Genève, Suisse (1970).
- /61/ "Mass Spectrometry and Organic Analysis-XIII. The Mass Spectra of Doubly Unsaturated Carbonyl Compounds", A.F. Thomas, B. Willhalm and Regula Müller, Org. Mass Spectrom. 2, 223 (1969).
- /62/ "Mise à Jour de Bibliothèques de Programmes (MACROSIM)", Code K409, Bibliothèque de Programmes, Université de Genève, Suisse (1970).

## Appendix A.

This review article was completed at the request of the editors of "Topics in Current Chemistry". Therein was presented some of the original work done on the ION GENERATOR program.

**Fortschritte der chemischen Forschung**  
**Topics in Current Chemistry**

*Reprint from Volume 39*

*Herausgeber: A. Davison, M. J. S. Dewar, K. Hafner, E. Heilbronner*  
*U. Hofmann, K. Niedenzu, Kl. Schäfer, G. Wittig*  
*Schriftleitung: F. Boschke*

Springer-Verlag Berlin Heidelberg New York 1973

Printed in Germany · Not in trade

---

A. B. Delfino and A. Buchs

**Mass Spectra and Computers**

## Mass Spectra and Computers

**Dr. Allan B. Delfino and Prof. Dr. Armand Buchs**

Department of Physical Chemistry, University of Geneva, Switzerland

### Contents

|                                                                     |     |
|---------------------------------------------------------------------|-----|
| I. The Learning Machine Approach .....                              | 110 |
| II. The Deduction Programming Approach .....                        | 112 |
| III. The Heuristic Programming Approach .....                       | 116 |
| A. The Meta-Dendral Project (MDP) .....                             | 125 |
| B. The Ion Generator Program (IG) .....                             | 125 |
| 1. Simple $\alpha$ -Cleavages Yielding $m/e$ 239 and $m/e$ 59 ..... | 131 |
| 2. McLafferty Rearrangement Ions .....                              | 132 |
| C. Future Work with the Ion Generator Program .....                 | 135 |
| IV. References .....                                                | 136 |

A. B. Delfino and A. Buchs

The mass spectrometrists see the digital computer in quite a different light from the theoretical chemist. In mass spectrometry the computer is not generally used to carry out approximations by numerical techniques which vary the parameters. It is mainly cast in the role of mechanical controller, data acquirer and processor, file searcher and as resolver of structures from mass spectra of unknown compounds.

Thus, data acquisition and the use of computers to monitor mass spectrometers, in gas chromatography/mass spectrometry combinations for example, will not be covered in this chapter. Excellent review and original papers have been published in this field <sup>1-12</sup>.

There are four general approaches to the computer-aided analysis of mass spectral data for structure determination. The first method, the use of file-searching techniques to correlate the mass spectrum of an unknown compound with reference spectra, really represents only a portion of the larger field of information storage and retrieval. Although file searching may be, at least for the times being, the most rewarding method available to the chemist for a rapid structure determination from spectral data, it will not be discussed here. For the status of this field of research the papers published by Hertz *et al.* <sup>13</sup>) and by Erni and Clerc <sup>14</sup>) make excellent reading.

The present chapter will be mainly devoted to the three other approaches for structure determination, *e. g.* the Learning Machine Approach (LMA), the Deduction Programming Approach (DPA) and the Heuristic Programming Approach (HPA). Much emphasis will be put on the artificial intelligence approach where the computer is used as a symbol manipulator. The last section will introduce a new departure in the use of a computer for mass spectrometry, the area of theory formation and the proposal of fragmentation mechanisms.

## I. The Learning Machine Approach

The LMA work of Jurs *et al.* <sup>15-23</sup>) in the field of mass spectrometry is important in applying the computer to structure identification. This work demonstrates what can be learned by the machine via a set of pattern dichotomizers or "binary pattern classifiers". The name binary pattern classifier comes from the fact that the pattern, in this case a mass spectrum, is placed in one of two categories. This name also implies that a classifier could, in fact, classify a pattern into one of many categories.

For our purposes, a binary pattern classifier is a computer program which is used to give a true or a false output when presented with a given mass spectrum. Thus the following question could be answered by either true or false:

Is oxygen present, in any functional form, in the substance which gave rise to this mass spectrum?

It was found that, by utilizing a sequence of such questions, empirical formulae could be determined from low-resolution mass spectra <sup>15</sup>. It took a series of 26 questions to determine an empirical formula of  $C_7H_{16}O_3N_2$  or less. The method is not error-free, but the error can be as low as 5%.

The work done with the LMA approach in the area of infrared spectroscopy, melting points and boiling points in conjunction with mass spectrometry <sup>18,19</sup> will not be dealt with in any detail here.

A pattern classifier vector,  $V$ , is used to generate the true or false response by forming a dot product between itself and  $S$ , the mass spectrum vector.

$$c = V \cdot S$$

Experimental results showed that the best way to interpret the scalar result  $c$  was as shown below:

$$\text{True} : c \geq + 50$$

$$\text{False} : c \leq - 50$$

$$\text{No answer} : - 50 < c < + 50$$

A result in the 'No answer' range of  $c$ , the so-called 'dead zone' range, indicates the inability of  $V$  to dichotomize the pattern in terms of the question for which  $V$  was designed.

The components of  $V$  are derived by a process of 'machine learning' which is called training. In order to train the vector  $V$ , a set of randomly selected mass spectra, called the training set, is chosen. The number of components of the mass spectrum vector  $S$  has to be one larger than the greatest mass found in any of the spectra forming the training set. If, for example, 178 is the largest  $m/e$  value in the training set of spectra, all spectra will become vectors of dimension 179. The intensity of  $m/e$   $i$  occupies the component  $S_i$ . The mass numbers are only implied by the existence of a component of  $S$  for each nominal mass. In our example, the position 179 is set to a value of 1 in all mass spectrum vectors of the training set, so that the 179th position of  $V$  will result in the position constant of the 178-dimension space function, where the decision plane, *i. e.* the boundary hypersurface between the true and the false hypervolumes, will have the orientation described by the 178 first components of  $V$ . Various methods were studied to find the best way to transform the spectra for the training process. The best results were obtained when the spectra were modified in such a way that the intensities were transformed to the square root of the per cent of the base peak intensity. The dimensionality of the vectors was found to be

A. B. Delfino and A. Buchs

reducible by various techniques such as graph theory <sup>22)</sup>, and double training for  $V$  <sup>20)</sup>.

The program used for training, the flow diagram for which is given by Jurs *et al.* <sup>15)</sup>, follows an error-correction procedure to arrive at acceptable values for the components of  $V$ . Initially all the  $V_i$  components have  $+1$  (or  $-1$ ) as their value. An initial dot product is then formed with the first mass spectrum in the training set. If the value of the dot product is outside the dead zone on the correct side, the response to the question, for example whether oxygen is present in the empirical formula, is deemed to be correct, and the next spectrum is processed. Whenever a spectrum is wrongly classified, a correction is made to  $V$ : a  $k$  multiple of the vector  $S$  is added to the old vector  $V$  in order to form a new vector  $V'$ .

$$V' = V + kS$$

The value of  $k$  is chosen so as to put the value of the dot product,  $V' \cdot S$ , on the correct side of the dead zone. Thus, this value is given by:

$$k = \frac{2}{|V|} (\pm M - c)$$

where  $M$  = size of the dead zone and  $c$  = original dot product of the incorrect answer.

New cycles through the complete training set are performed until no further corrections are necessary. The value of  $V$  will be found acceptable when all the spectra of the training set are classified correctly. After training is completed, predictive ability can be tested. Jurs <sup>21)</sup> reports, for example, 98% predictive ability on a set of 300 mass spectra (89 from oxygen-containing compounds and 211 from substances containing no oxygen). The vector used on the 300 mass spectra had been trained on a different set of 300 mass spectra, 84 of them originating from oxygen-containing compounds. In order to achieve 100% recognition on the training set spectra, 375 feedback corrections were necessary. A similar process was unfortunately not carried out on the 26 vectors required to infer empirical formulae from low-resolution mass spectra. Such an experiment would, no doubt, be of great value in determining the potential of the LMA approach.

## II. The Deduction Programming Approach

This section concerns the use of non-heuristic programs to deduce structural information from mass spectra. Automated systems were proposed as early as 1966 for compound-type identification <sup>24,25)</sup> and as an aid in the inter-

pretation of mass spectra <sup>26)</sup>. Among them are the well-known polypeptide sequencing programs <sup>27-29)</sup>, the dialogue program designed by Biemann and Fennessey <sup>30)</sup> and the subsequent man-computer interaction program for resolving complex chemical compounds <sup>31)</sup>. The discussion will be restricted to the recent work we consider the most general and the most relevant.

Crawford and Morrison <sup>32)</sup> have designed a general program to extract structural features from low-resolution mass spectra rather as a chemist does it. The program uses various interrogation schemes to yield successively the molecular mass, the presence of functional groups, and finally the molecular skeleton. Its success varies according to the complexity of the structure; a chemist might also have varying results if he was compelled to work only with low-resolution mass spectra as data. But the computer-controlled GC/MS systems, produce such enormous amounts of data that they must be processed by a computer. If it does not determine exact structures, at least the computer should reduce the possibilities, so that the mass spectrometrists can use this information to work out structures more quickly.

The main program deals with input and output requests and of various requests for structure determination. It coordinates the reading and the preprocessing of the mass spectrum, asks a subprogram for the molecular weight and the functional groups most likely to be present, calls on routines which specialize in various classes of mass spectra, calls for a nearest-neighbor table <sup>33)</sup> if more specific information has not yielded reliable answers, and lastly asks *that output be made* of the results of all the other dispatches made previously.

The program for determining the most probable group was based on earlier work of Crawford and Morrison <sup>34)</sup> in which average mass spectra, representative of various functional groups, were presented as points on a hypersphere. This program recognizes 12 classes of compounds:

Aromatics,  
esters,  
ethers,  
acids,  
ketones,  
aldehydes,  
alkenes  
alkanes,  
alcohols,  
cycloalkanes,  
dienes and  
amines.

A. B. Delfino and A. Buchs

The molecular-weight routine assumes the highest  $m/e$  value to be near the molecular weight. If the molecular ion is not in the spectrum, the fact that it was not observed will be printed. The routine tries to distinguish between isotope peaks and the molecular-weight peak; if it decides that a certain peak is a likely candidate for the molecular ion, it checks whether there are any forbidden peaks in the mass range  $M-3$  to  $M-13$ . A consistency check based on the greater stability of even-electron over odd-electron ions is also performed.

The empirical-formula routine uses the isotopic abundances and the inferred molecular mass and functional groups to arrive at upper limits for carbon content and degree of unsaturation.

If the inferred empirical formula is descriptive of a certain class of compounds whose fragmentation is known to the system, the special class routine for this type of compounds is brought in. If the class is not well known to the program, a general interrogation routine is enlisted for further processing.

The general interrogation routine does the 'inspired guesswork' of the system for classes which have no special class routine; it is based on the diagnostic key fragments proposed by McLafferty <sup>35</sup>. Associated with each key fragment is a piece of basic structure and some peak characteristics. Once the characteristics of the peak have been sensed, the basic structure is remembered for integration into the final result later on, this result is based on a probability of correctness calculated for each inferred basic structure. Finally, a structure-drawing routine prints the most probable structure.

To place their work in proper perspective, Crawford and Morrison compared the outcome of the program with the abilities of fourth-year students. The humans and the computer were about equal in their capacity to interpret the given mass spectra. The real significance of this work is that it is an *ab initio* approach, based not only on theory, but also on empirical evidence drawn from large numbers of reference mass spectra.

A method able to classify compounds into various functional categories from their low-resolution mass spectra, similar to the method used by Crawford and Morrison, has been recently proposed by Smith <sup>36</sup>. A large set of reference spectra is reduced to a correlation set of 'ion-series spectra'. Smith shows that each class of compounds that is available in an archive can be characterized by a typical ion-series spectrum. To determine the class of compounds to which an unknown substance should be assigned, the intensities of the signals found in its low-resolution mass spectrum are summed in the 14 different ion series calculated according to a formula similar to that used by Crawford and Morrison. Smith uses the following equation:

$$S_m = \frac{\sum_n I(30 + m + 14n)}{\sum I(j)}$$

where  $I(j)$  is the intensity of mass  $j$ ,  $m = 1$  to  $14$ ,  $n = 0, 1, 2, \dots$  and where  $S_m$  represents the per cent contribution of ion series  $m$  to the total ion current.

This summation yields the ion-series spectrum for that compound, which is then matched, either manually or automatically, with the ion-series spectra of each class of compounds known to the system. Applications to environmental chemistry and to the analysis of complex geochemical samples seem to have been promising.

Although most of our discussion concerns programs that process low-resolution mass spectra, this does not indicate a bias on our part for low-resolution mass spectrometry. It is attributable rather to the fact that high-resolution work requires more in the way of instrumentation, so that low-resolution spectrometers are much more common. Hence people tend to focus their computing research toward the greater audience of low-resolution mass spectrometrists.

The work of Venkataraghavan, McLafferty and van Lears <sup>37)</sup> is concerned with problems associated with structure determination on the basis of data acquired with high-resolution instruments. These authors present a generalized program to find structures from the elemental compositions of fragment ions. The program, as described in the publication mentioned above, could operate on ketones, esters, alcohols, amines and keto-alcohols containing saturated acyclic hydrocarbon moieties. The first part of the program tries to identify the general nature of the compound and its functional groups. Next specific subroutines search for known fragmentation pathways, typical of the compound class postulated. Much work is involved to identify a molecular ion. The candidate ion must contain the highest number of atoms of each element present in any ion of the spectrum, and its unsaturation degree must be zero or some positive integer. A consistency check is then performed by using the masses of the neutral fragments lost from the candidate molecular ion to yield the major fragment ions at the high-mass end of the spectrum. The number of occurrences of commonly lost fragments is weighed against those losses which are highly improbable. When this test rejects a postulated molecular ion, the program proposes candidates of higher molecular weight in a way similar to that proposed by Biemann and McMurray <sup>38)</sup>. Losses of neutral fragments are further used to indicate the types of structures present, which are printed out by the computer.

The next stage of the program processes the information contained in characteristic ion series. Searches are conducted for homologous as well as for non-homologous important ion series. With that information, the system then attempts to deduce the most probable compound classes and then to postulate structures within the chosen classes. Using character-

A. B. Delfino and A. Buchs

istic fragmentation pathways of a class, the program works its way back to a structure.

The power that can be built into routines for structure determinations restricted to particular classes of compounds is clear from the later discussion concerning the Inference Maker program. If the programs of the system proposed by Venkataraghavan *et al.* can be made as powerful as the Inference Maker, it could become the most effective of those reported to date.

### III. The Heuristic Programming Approach

One of the most successful approaches to structure identification by means of computers was achieved at Stanford University with the *Heuristic Dendral Program*, based on the dendritic algorithm of Lederberg<sup>39</sup>. The Heuristic Dendral Program (HDP), as it was before 1970, is explained in various publications<sup>40-42</sup>, and summarized in Fig. 1. The data, a mass

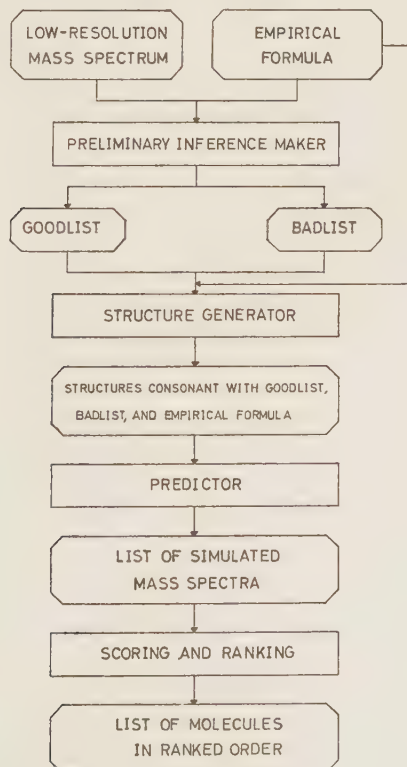


Fig. 1. Flow diagram of HDP

spectrum and an empirical formula, are given to the multi-phase HDP, and the planning phase, called the Preliminary Inference Maker (PIM), ascertains the presence or absence of the functional groups known to the program as a whole. The 'heuristics' used in this phase are similar to the rules that a mass spectrometrists applies to deduce the presence or absence of functional groups. These rules were given to the program in the form of tables. The results of the PIM phase are sorted into two parts: functional groups possibly present are put into a table called Goodlist and other functional groups known to be absent are put into Badlist. These tables are used as input to the next phase of HDP, called the Structure Generator (SG).

The SG uses the empirical formula along with Goodlist and Badlist to generate *exhaustively* but without redundancy a list of candidate structures for the form of the substance which gave rise to the mass spectrum. These candidates, consonant with the empirical formula, Goodlist and Badlist are then given as input to the third phase of HDP, the spectral simulator, called the Predictor.

After having generated characteristic mass spectral features from each candidate structure, the final phase of HDP, appropriately called the Scoring phase, compares the simulated mass spectrum and the given mass spectrum and scores the hypothetical structures according to the fit of the calculated mass spectrum to the actual mass spectrum. The candidate structures in order of descending score represent the final output of the HDP. If the program were correctly designed, the correct structure would have the highest score.

The HDP, operating as above, was tested with aliphatic ketones <sup>41)</sup> and with aliphatic ethers <sup>42)</sup>. This work underscored the lack of theory in the PIM phase. It was possible to generate unnecessarily large numbers of structures at great cost, only to have them ranked low or equal by the Predictor phase. As the size of the molecule increased, the SG phase was producing structures for the most part indistinguishable by their mass spectral behavior. This situation led to the evolution of the PIM program into the single-phase Inference Maker (IM) program which was tested on the mass spectra of saturated acyclic amines. Following the construction of the single-phase program with its repertoire of general subprogram units to determine amine structures <sup>43)</sup>, further and more powerful generalizations were made to give rise to the most recent program, summarized below, and described in one of the Dendral Project papers <sup>44)</sup>.

The current version of the Inference Maker program, which is written in the LISP programming language and also in FORTRAN, here in Geneva, can deduce the structure of any compound belonging to the class which is encompassed by the empirical formula  $C_nH_{2n+v}X$  ( $X=O, S$  or

A. B. Delfino and A. Buchs

N,  $v$  = valence of X). This class of Saturated Acyclic Monofunctional (SAM) derivatives includes amines, alcohols, thiols, ethers and thioethers. The emphasis was on keeping the program units as general as possible but tuning their 'heuristics' to ensure drastic pruning, yet never allowing the correct structure to be absent from the final output of candidates.

If the IM can deduce the nature of the heteroatom along with its molecular weight, the empirical formula can be calculated for this class of compounds, because they have a zero unsaturation degree. So, if the IM can filter out the mass spectra which do not originate from SAM compounds, the empirical formula can be precisely determined. Search methods for heteroatom and molecular weight are explained below. The rules governing the fragmentation modes of SAM molecular ions are general enough yet precise enough to enable the program to calculate the theoretical parameters required to elucidate structures of this class, thus relieving the chemist of this boring and time-consuming task.

The following discussion of how the IM processes mass spectral data is summarized in Fig. 2. At the outset, the filter program removes from the given mass spectrum all the ions belonging to the series of ions which could be used later on in the process of structure determination. These are the  $\alpha$ -cleavage ion series, starting from all the masses of the general formula  $\text{CH}_2\text{XH}_{v-1}$ , *i.e.* from 30, 31 and 47 where X is a nitrogen, an oxygen and a sulfur atom, respectively. The three alkyl ion series  $\text{C}_n\text{H}_{2n+1}$ ,  $\text{C}_n\text{H}_{2n}$  and  $\text{C}_n\text{H}_{2n-1}$  are also removed, because of the role they play in the spectra of oxygen- and sulfur-containing SAM compounds. If this 'reduced' mass spectrum still contains any intense ions (10%), or if the remaining ions show, on the average, an abundance greater than 3%, the filter cancels further processing of the mass spectrum. A spectrum which is reduced enough to qualify as a possible SAM spectrum is then ready to undergo the heteroatom-hypothesis process.

Using the  $\alpha$ -cleavage ion series, the IM sums for each heteroatom the total intensity found in the corresponding  $\alpha$ -series until the largest  $m/e$  value present in the spectrum is smaller than a given member of the homologous series. This intensity sum becomes the heteroatom's score. The heteroatom with the highest score is the one most likely to be contained in the empirical formula. The other heteroatoms are ranked in descending order of score and held for later processing, if required. If the first heteroatom does not give rise to a consistent structural output, the program might then be able to find its way out of a wrong choice of heteroatom due to some unforeseen weakness in the scoring method, or to some error in the mass spectrum. The rest of the program is more stringent in its requirements. Hence, an improper heteroatom should give rise to no output whatsoever. In fact, we had the program make bad first choices on purpose.

Mass Spectra and Computers

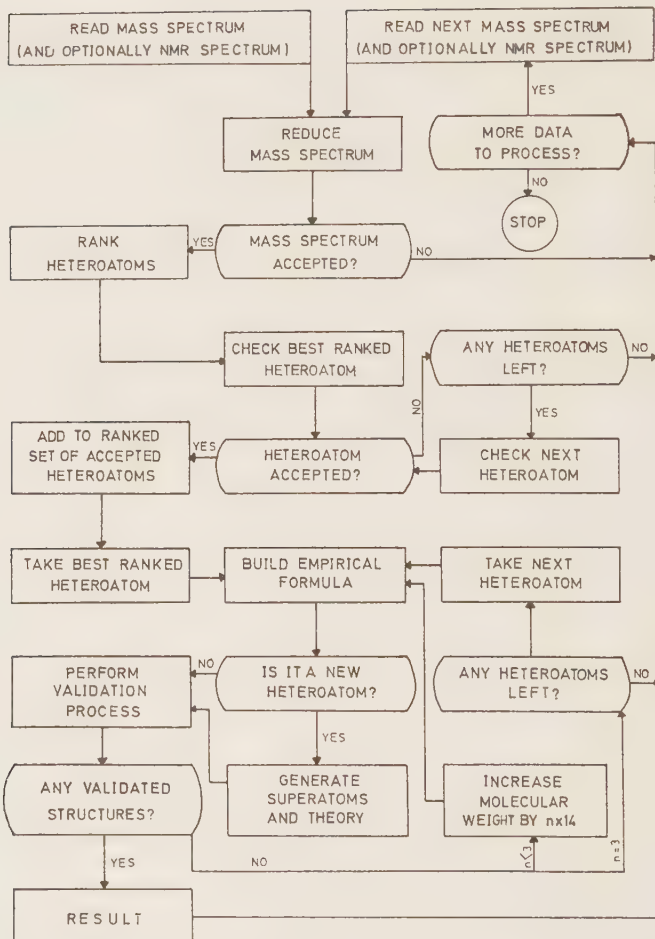


Fig. 2. Flow diagram of the IM program

It never found a structure corresponding to a false heteroatom and always went on with the process of validation until the correct heteroatom led to proper structure determination.

Intensities found in the alkyl ion series are also used to validate a heteroatom choice. Since nitrogen provides a better locus for the positive charge than either sulfur or oxygen, amine mass spectra carry a large proportion of the total ion current in the  $\alpha$ -cleavage ion series, showing only a few

A. B. Delfino and A. Buchs

hydrocarbon-type ions. So a heteroatom will be kept for further processing only if the distribution of the ion current between the  $\alpha$ -cleavage and the alkyl ion series is in accord with its charge retention capacity.

Once a heteroatom is chosen, a minimum molecular weight is inferred on the basis of the high mass range of the spectrum and the knowledge of the heteroatom. This then gives rise to an empirical formula. As with the heteroatom phase, the molecular weight determination phase is also a hypothesis stage, with the minimum value ranked as most plausible, and the next two higher homologs ranked after it. Failure to produce a structural output with one molecular weight causes the next higher homolog to be tried, until three values have been considered. Failure with all three values tried as molecular weights will cause a return to the heteroatom-hypothesis stage explained above. If no heteroatom, with any empirical formula, gives rise to an output, the program gives up and classifies the spectrum as not originating from a SAM molecule.

Having decided upon the empirical formula, the program then builds up a complete and nonredundant set of structural subunits called 'superatoms'. These superatoms contain the heteroatom and all the carbon atoms in the  $\alpha$  positions. For each heteroatom, the complete set of superatoms represents every possible number of  $\alpha$ -carbon atoms in accord with the valence of the heteroatom; and for each of these  $\alpha$ -carbon atoms every type of substitution degree is represented. These superatoms are generated by the IM as combinations of the four letters M, P, S and T which stand for a methyl, a primary, a secondary and a tertiary carbon atom respectively. The defined canonical order of the symbols ( $M < P < S < T$ ) demands that, in any sequence of symbols which will denote a superatom, any symbol must be to the right of a symbol which has a lower canonical value; thus, SPM is the only valid form from the set of the structurally equivalent forms: MPS, MSP, PSM, SMP, PMS, and SPM. For divalent heteroatoms only 12 groupings can be generated (P, S, T, PM, PP, SM, SP, SS, TM, TP, TS, TT), but for the trivalent nitrogen there are 31 different groupings which can be built. They are the twelve mentioned above, plus all the three symbol representing the tertiary amines: PMM, PPM, PPP, SMM, SPM, SPP, SSM, SSP, SSS, TMM, TPM, TPP, TSM, TSP, TSS, TTM, TTP, TTS and TTT. To specify the nature of the inferred heteroatom, a prefix is placed in front of the  $\alpha$ -substitution symbol sequence. This completes the name of the superatom. These prefixes are AM, EA and TH for N, O and S respectively. Illustrated in Fig. 3 are some examples of superatom names along with the structures they represent.

The  $\alpha$ -carbon atom environment, implied by the name of the superatoms, enables the IM to decide which mode of fragmentation would be exhibited by compounds with a particular superatom as the central structural subunit, after the alkyl radicals have been specified. Some of the rules

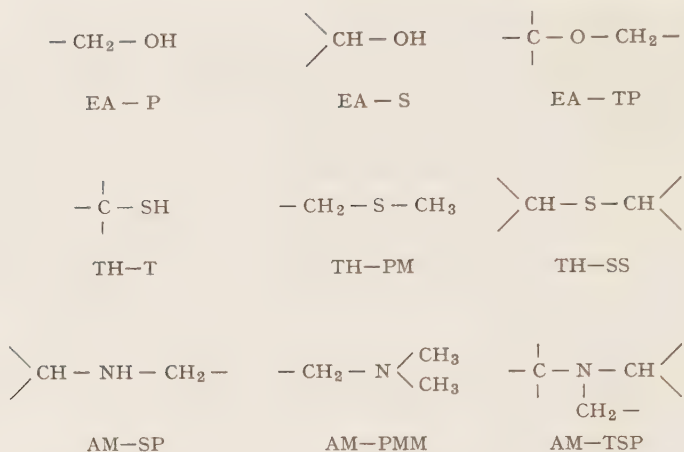


Fig. 3. Structures of some superatoms and their corresponding names

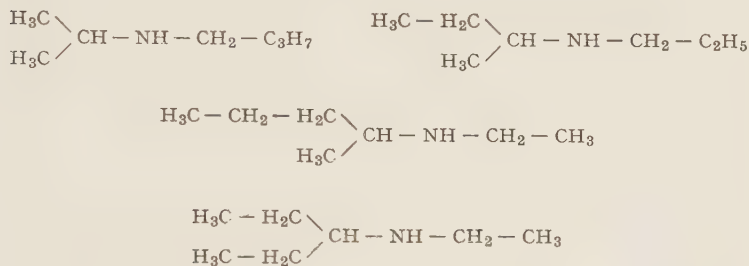
which govern the fragmentation of a molecule are specific to the superatom contained in that molecule, others are generally applicable to all the superatoms of a given heteroatom. The marked tendency of SAM molecular ions to fragment via  $\alpha$ -cleavage is an example of a generally applicable rule.

The validation process begins with only those superatoms which can be contained in the empirical formula when a methyl group is attached to every free valence of the superatom. Superatoms requiring more carbon atoms than are available in the empirical formula are rejected out of hand. Early in the validation process an NMR spectrum can be optionally utilized to reject those superatoms which require more terminal methyl groups than are exhibited in the NMR spectrum. The number of heteroatom-bound methyl groups and the number of heteroatom-bound hydrogen atoms, deduced from the NMR spectrum, have also to be consistent with the structure of a superatom before it is retained for further processing.

There comes a time in the validation process when the surviving superatoms will be transformed into general molecules by the specification of all possible sets of alkyl radicals, of unknown inner structure, which have to be attached to the free valences of the superatoms. To which of the free valences each possible alkyl radical may belong is not important at this point. The atoms of the superatom under consideration are subtracted from the empirical formula so far inferred, leaving the number of carbon atoms which must be distributed in all possible ways among the free valences of the superatom. For example, if the inferred elemental composi-

A. B. Delfino and A. Buchs

tion was  $C_7H_{17}N$ , superatom AM-SP would be extended to the four structures shown below (1):



1

Each of these four general structures is then tested further, and the final output shows those which have passed all the tests related to  $\alpha$ -cleavage, rearrangement ions, alkyl ions, etc. A complete description of the rules used to treat these molecules as unknowns can be found in one of our previous publications <sup>44</sup>.

Diagram 1 shows the results printed by the IM when the *mass spectrum of 3-octanol* is processed without the help of an NMR spectrum. Once the program has inferred oxygen as the most plausible heteroatom, and  $C_8H_{18}O$

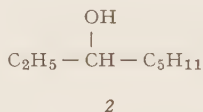
Diagram 1. Output of the IM program with octane-3-ol as an unknown

---

|                                                                                                                                                                                                                                                                                                                                                                                                |           |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| ACTUAL MASS SPECTRUM = ((26 . 3) (27 . 31) (28 . 7) (29 . 35) (30 . 2) (31 . 32) (36 . 1) (38 . 1) (39 . 13) (40 . 2) (41 . 38) (42 . 7) (43 . 20) (44 . 9) (45 . 7) (51 . 1) (53 . 2) (54 . 1) (55 . 48) (56 . 1) (57 . 14) (58 . 9) (59 . 100) (60 . 4) (67 . 1) (68 . 1) (69 . 4) (70 . 3) (71 . 2) (73 . 1) (81 . 1) (82 . 1) (83 . 42) (84 . 4) (100 . 1) (101 . 16) (102 . 1) (112 . 2)) |           |
| RESIDUE MASS SPECTRUM = ((51 . 1) (53 . 2) (54 . 1) (67 . 1) (68 . 1) (81 . 1) (82 . 1))                                                                                                                                                                                                                                                                                                       |           |
| MASS SPECTRUM CORRECTED FOR 13-C = ((26 . 3) (27 . 31) (28 . 7) (29 . 35) (31 . 32) (36 . 1) (39 . 13) (40 . 2) (41 . 38) (42 . 6) (43 . 20) (44 . 8) (45 . 7) (51 . 1) (53 . 2) (54 . 1) (55 . 48) (57 . 14) (58 . 8) (59 . 100) (67 . 1) (68 . 1) (69 . 4) (70 . 3) (71 . 2) (73 . 1) (81 . 1) (83 . 42) (84 . 2) (100 . 1) (101 . 16) (112 . 2))                                            |           |
| WAS A NMR SPECTRUM AVAILABLE                                                                                                                                                                                                                                                                                                                                                                   | NO        |
| INFERRED MOLECULAR WEIGHT                                                                                                                                                                                                                                                                                                                                                                      | 130       |
| INFERRED EMPIRICAL FORMULA                                                                                                                                                                                                                                                                                                                                                                     | C8H18O    |
| GENERAL MOLECULES INFERRED:                                                                                                                                                                                                                                                                                                                                                                    |           |
| *EA-S*(C5H11,C2H5)                                                                                                                                                                                                                                                                                                                                                                             | 8 ISOMERS |

---

as the empirical formula, the total search space contains 171 *a priori* valid structures, *i.e.* all the isomers having that empirical formula. Despite the fact that no molecular ion is present in the mass spectrum, the correct empirical formula is inferred by using the postulated minimum molecular weight value. If an NMR spectrum is used in conjunction with the mass spectrum, the presence of only two carbon-bound methyl groups and the absence of any oxygen-bound methyl radical allows the IM to reject straight away any superatoms with more than two free valences, and those superatoms with the symbol M in the name. This reduces the exhaustive, non-redundant set of twelve superatoms to only three. The three which survive the NMR tests are EA-P, EA-PP and EA-S, the last being the correct one. Moreover, the presence of only two methyl groups reduces the number of isomers for the general molecule EA-S-(C<sub>2</sub>H<sub>5</sub>, C<sub>5</sub>H<sub>11</sub>), shown in Diagram 1 and given below (2), to one of the eight isomers which are possible the structure of the C<sub>5</sub>H<sub>11</sub> radical is unknown.



The IM program has been tested with 210 mass spectra, including amines, alcohols, ethers, thiols and thioethers. The correct structure was always included in the output and, even for large structures, the correct structure was often the only one proposed as an answer. Table 1 shows some results with various SAM compounds. Complete lists of the performance of the IM program are given in two previously mentioned publications <sup>43,44</sup>. The program is able to process a mass spectrum in about 3 seconds on an IBM 360, Model 65 computer. When an NMR spectrum is supplied along with the mass spectrum, the processing is even faster, because, as was explained above, superatoms can be rejected early on in the process, by means of simple tests. From the results so far attained, we feel that the IM program can determine SAM structures better than an experienced mass spectrometrists. The performance which was achieved should encourage the development of other programs for special classes of compounds.

Besides the heuristic programming approach oriented toward structure determination, there are two other challenging areas where heuristic programming is applied to mass spectrometry. Both these areas involve the use of the computer mainly as a symbol manipulator. Automated formation of a theory <sup>45</sup> and computer rationalization of mass spectral fragmentation mechanisms <sup>46</sup> are two topics of ongoing research that might eventually lead to a better formalization of the rules governing fragmentation under electron impact.

A. B. Delfino and A. Buchs

Table 1. Some IM program results

| <i>n</i>          | $C_nH_{2n+v}X$ compounds                   | Number possible<br>$C_nH_{2n+v}X$<br>isomers | Number of inferred<br>isomers<br><br>A <sup>a)</sup> | B <sup>b)</sup> |
|-------------------|--------------------------------------------|----------------------------------------------|------------------------------------------------------|-----------------|
| <i>Amines</i>     |                                            |                                              |                                                      |                 |
| 6                 | Tri-ethyl                                  | 39                                           | 2                                                    | 1               |
| 8                 | 3-Octyl                                    | 211                                          | 26                                                   | 2               |
| 9                 | Tri- <i>n</i> -propyl                      | 507                                          | 2                                                    | 1               |
| 10                | Di-isopentyl                               | 1.238                                        | 109                                                  | 16              |
| 11                | <i>n</i> -Undecyl                          | 3.057                                        | 507                                                  | 1               |
| 14                | Di- <i>n</i> -heptyl                       | 48.865                                       | 646                                                  | 1               |
| 15                | Tri- <i>n</i> -pentyl                      | 124.906                                      | 40                                                   | 1               |
| 17                | N-methyl-bis-2-ethylhexyl                  | 830.219                                      | 2.340                                                | 24              |
| 18                | N-methyl- <i>n</i> -octyl- <i>n</i> -nonyl | 2.156.010                                    | 15.978                                               | 1               |
| 20                | N,N-dimethyl- <i>n</i> -octadecyl          | 14.715.813                                   | 1.284.792                                            | 1               |
| <i>Ethers</i>     |                                            |                                              |                                                      |                 |
| 6                 | Ethyl sec-butyl                            | 32                                           | 2                                                    | 2               |
| 6                 | Di-isopropyl                               | 32                                           | 1                                                    | 1               |
| 7                 | Isopropyl sec-butyl                        | 72                                           | 3                                                    | 2               |
| 9                 | <i>n</i> -Butyl <i>n</i> -pentyl           | 405                                          | 8                                                    | 1               |
| 20                | Di- <i>n</i> -decyl                        | 11.428.365                                   | 22.366                                               | 1               |
| <i>Alcohols</i>   |                                            |                                              |                                                      |                 |
| 7                 | 3-Ethyl-3-pentyl                           | 72                                           | 1                                                    | 1               |
| 8                 | 2,3,4-trimethyl-3-pentyl                   | 171                                          | 3                                                    | 1               |
| 10                | 6-Ethyl-3-octyl                            | 989                                          | 39                                                   | 9               |
| 12                | 2-Butyl-1-octyl                            | 6.045                                        | 1.238                                                | 25              |
| 14                | 3-Tetradecyl                               | 38.322                                       | 1.238                                                | 1               |
| <i>Thioethers</i> |                                            |                                              |                                                      |                 |
| 7                 | Isopropyl sec-butyl                        | 72                                           | 4                                                    | 3               |
| 8                 | Ethyl <i>n</i> -hexyl                      | 171                                          | 8                                                    | 1               |
| 8                 | Di-sec-butyl                               | 171                                          | 3                                                    | 1               |
| 10                | Di- <i>n</i> -pentyl                       | 989                                          | 12                                                   | 1               |
| 14                | Di- <i>n</i> -heptyl                       | 38.322                                       | 153                                                  | 1               |
| <i>Thiols</i>     |                                            |                                              |                                                      |                 |
| 6                 | 1-Hexyl                                    | 32                                           | 8                                                    | 1               |
| 6                 | 2-Hexyl                                    | 32                                           | 12                                                   | 5               |
| 8                 | 1-Octyl                                    | 171                                          | 39                                                   | 1               |
| 10                | 1-Decyl                                    | 989                                          | 211                                                  | 1               |
| 12                | 1-Dodecyl                                  | 6.045                                        | 1.238                                                | 1               |

<sup>a)</sup> A = Inferred isomers when only mass spectrometry is used.

<sup>b)</sup> B = Inferred isomers when the number of methyl radicals is known from NMR data.

### A. The Meta-Dendral Project (MDP)

The MDP being undertaken at Stanford University seeks to write a program which, given structures and corresponding mass spectra, will be able to discover the parts of a unifying theory of mass spectrometry. The program starts with a minimum of assumptions. When enough of the parts have been identified, it will unify them into a description of how molecular ions fragment under electron impact. The overall strategy of the MDP program can be divided into three parts:

1. Given a set of operations called, for example, Breakbond, Migrategroup, etc., etc., apply them in all possible singlets, doublets, triplets, up to some optional multiplet, to the given input structure. The data are the low-resolution mass spectrum and the corresponding structure. The output will consist of ionic structures and the sequence of processes that could give rise to them. These outputs will become inputs to the second part of the program. This step in no way uses an existing theory, which would be a form of question-begging for a program such as this. No information exists about charge loci, radical site triggering, or new bond formations as tools for this heuristic search.
2. Given the output of stage 1 for many related compounds, the program will generalize the results to find which sequences of processes exist in all the related compounds. Then it will search the retained ionic structures, for the common factor that allowed the process to occur. This distilled essence, along with the process set, becomes input to the final stage of the MDP program.
3. The output of stage 2 must finally be integrated into the existing theory. At the outset, this will, of course, be zero. But as soon as two structural essences and two process sets exist, it must be determined whether the second is just a corollary of the first, or vice versa, or whether they are disjoint situations with their associated actions. The theory will then take the form of these so-called situation-action rules <sup>45)</sup>. The situations are the structural essences and the actions are the process sets.

Since the project has only just started, no definite chemical results have yet been published. But the explanation of the program, in a paper addressed to computer scientists <sup>45)</sup>, gives hope that the project will soon bear some encouraging results for chemistry.

### B. The Ion Generator Program (IG)

The approach taken at the University of Geneva <sup>46)</sup> addresses quite a different problem. We have devised a heuristic search program, the *Ion Generator* program, to simulate the formation of ions in a mass spectrometer.

A. B. Delfino and A. Buchs

It is based on the well-known method of electron bookkeeping, widely used by mass spectrometrists to explain the formation of a particular ion from its precursor ions. The IG generates, exhaustively, all motivated sequences of electron book-keeping processes, and heuristically selects those sequences, called mechanisms, which might help to formalize mass spectral fragmentations. It is in its exhaustiveness that the program differs from the chemist. The chemist is prone to stop after he finds one reasonable mechanism; the program will find all of the reasonable mechanisms.

The input to the IG program is a structure and its mass spectrum. The output is a list of ions with their masses and structures, and with the mechanisms which gave rise to each ion generated. The output also includes, for each ion, the intensity with which it appears in the actual mass spectrum. We call these electron book-keeping processes primitive operations; a sequence of primitive operations constitutes a mechanism. For some of the generated ions, only one mechanism will be proposed; for other ions, many sound mechanisms may result. Some mechanisms will lead to the formation of ionic species which are absent from the mass spectrum. A mechanism which leads to an ion that was not present in the mass spectrum, and whose daughter ions are also missing, could be a candidate for rejection. Of course, the real test for mechanism rejection would be experiments with labeled compounds.

In generating molecular ions from molecules, and fragment ions from molecular ions, the program applies only the primitive operations which are *motivated* by certain structural environments that exist in the subject structure. It does not perform every primitive operation in every situation. The IG program rests on the assumption that the given motivations to perform the primitive operations are a sound basis on which to work. In its current implementation, the IG will perform the following primitive operations, given the proper structural situation:

- Ionization of the structure with charge localization.
- Homolysis of bonds in the  $\beta$  positions of an existing radical site.
- Bond formation by the pairing-up of two adjacent radical sites.
- Transfer of an atom or of a group of atoms to an existing radical site via cyclic transition states of various sizes.
- Ring closures between existing radical sites or between one existing radical site and another radical site created by a simultaneous homolysis.

It is envisaged that these primitive operations will be only a part of the complete set of operations used by the program in the future.

In order to generate sequences of primitive operations in an exhaustive and nonredundant way, the IG program first checks that the required structural motivations exist for all primitive operations in the molecular

ion. These motivated operations result in transformed structures. Some of these resultant structures will be fragment ions, but others will be rearranged forms having the same mass. The rearranged ions will undergo the same check for structural motivations for all the primitive operations. Hence, all possible sequences will be applied to the molecular ion to result in every possible daughter ion. The nonredundance results from the fact that only one motivated operation is done at one time on any of the intermediate structures.

At present, the program is restricted to generating only primary ions. However, nothing in the generation scheme prevents the program from going further; it is so restrained because of the amount of output that is generated. When more is learned about methods which can justifiably be applied to limit the amount of output, secondary and higher order ions will be generated also.

The motivations and the restrictions put on the primitive operations are summarized in Diagram 2, and will briefly be described below.

Diagram 2. Application rules of the primitive operations of the IG program

| Primitive operations               | Motivations                                                                                                                                   | Restrictions                                                                                                                                                                                                |
|------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ionization                         |                                                                                                                                               | Selected ionization threshold range                                                                                                                                                                         |
| Homolysis                          | Bond $\beta$ to a radical site: formation of a new bond                                                                                       | Bond was formed in a previous step of the ongoing mechanism                                                                                                                                                 |
| Makebond                           | Presence of two adjacent radical sites                                                                                                        | Bond was broken in a previous step of the ongoing mechanism                                                                                                                                                 |
| Transfer of R.<br>R = H or radical | Presence of a radical site. Appropriate size for a cyclic transition state                                                                    | No blatant strain in the transition state. Shifted bond must not have been formed in a previous step of the ongoing mechanism. Selected maximum number of successive transition states must not be exceeded |
| Ring closure                       | Two existing radical sites at proper distance. Intramolecular attack of a bond due to reactivity of the radical site. Formation of a new bond | Original broken bond must not have been formed in a previous step of the ongoing mechanism. Allowed sizes for the rings formed. No blatant strain in resulting ring                                         |

A. B. Delfino and A. Buchs

It is well known that the most successful explanation of the fragmentation which occurs under electron impact still relies on the concept of charge localization, just prior to fragmentation. This concept is accepted by most mass spectrometrists, and its importance is underlined in textbooks dealing with organic mass spectrometry. Our program has been designed to recognize that the site on which the positive charge is best accommodated is structure-dependent. Functional groups containing heteroatoms will be preferred as charge loci over carbon-carbon sigma bonds or carbon-carbon double bonds. For each different type of electrons ( $n$ ,  $\pi$ ,  $\sigma$ ), and for each environment, an ionization threshold value has been semi-empirically chosen. Nitrogen is known to accommodate the positive charge better than sulfur or oxygen. So, in the sequence of ionization threshold values, ionization on a nitrogen atom will have priority over ionization on a sulfur atom or on an oxygen atom. Table 2 shows ionization threshold values selected for a number of structural environments. The user of the program is free to choose the complete range of ionization threshold values, or to

Table 2. Ionization threshold values of the IG program

| Structural environment | Ionization threshold value |
|------------------------|----------------------------|
| $>N-$                  | 1000                       |
| $>NH$                  | 950                        |
| $-NH_2$                | 900                        |
| $>C=S$                 | 850                        |
| $>C=O$                 | 800                        |
| $-S-$                  | 750                        |
| $-O-$                  | 700                        |
| $-SH$                  | 650                        |
| $-OH$                  | 600                        |
| $-C\equiv C-$          | 550                        |
| $>C=C<$                | 500                        |
| $>C<$                  | 450                        |
| $\geq CH$              | 400                        |
| $>CH_2$                | 350                        |
| $-CH_3$                | 300                        |

select a smaller range with optional upper and lower limits. If, in a structure containing a keto group, one wants ionization to occur only by removal of a lone-pair electron from the carbonylic oxygen atom, the ionization threshold value range can be chosen from 750 to 850. On the other hand, difunctional molecules can be ionized either on both heteroatoms in the same run or on each heteroatom separately in two successive runs. This affords an interesting possibility to study the additivity property of mass spectral

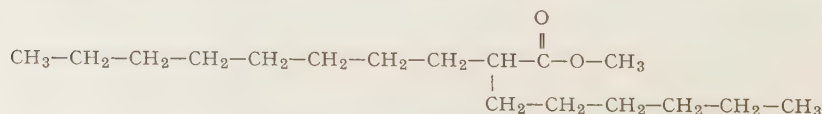
fragmentations. When a rather broad range of ionization threshold values is allowed, the program builds and fragments the molecular ions in the order of the priority scale shown in Table 2.

The only homolytic cleavages which are presently allowed to occur are those involving bonds which are in the  $\beta$  positions with respect to an existing radical site. The motivation to make a fission is the potential to form a new bond by radical pairing, either in the ionic or in the neutral species. This is known to be an energetically favorable action. Around a radical site, every bond which is in a  $\beta$  position is fissionable, unless it has been formed by the IG program in a previous operation of the ongoing process. Homolysis of  $\beta$  bonds will also be suppressed if the resulting intermediate or final species contains a number of radical sites exceeding the number which has been chosen by the user. The formation of a bond by radical pairing will occur if the resulting bond is not one which has been broken by the IG in a previous step.

Another way to form a new bond at a radical site is by intramolecular attack of a bond via a cyclic transition state. This leads to transfer of a hydrogen atom or of a group of atoms to the radical site, thus creating a new radical site able to trigger either homolysis or a new cyclic transition state. The user chooses the number of transition states allowed to occur in any sequence of primitive operations, as well as the sizes of the rings. Most of the time we allow only two or less successive cyclic transition states to occur in a mechanism. The available sizes of the transition states range from four-membered to eight-membered rings. The Ion Generator will not attempt to go through blatantly strained transition states. Another interesting feature of the IG is that it distinguishes between 'long-lived' and 'short-lived' transition states. When the first transition state applied is long-lived, the spatial conformation of the structure is retained until the second transition state has been completed. This allows two transition states to occur simultaneously, in a concerted way.

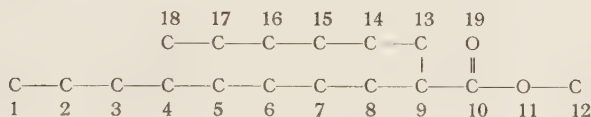
The program knows that intramolecular attack of a bond by a radical site does not always lead to transfer of an H atom or of a radical, but can be followed by cyclization with simultaneous elimination of the radical which was attached to the attacked bond; ring structures are thus obtained. An example of such a ring formation is given in Scheme 2 (see p. 134).

To illustrate the performance of the IG, some examples of mechanisms proposed for the formation of primary ions from compound 3 will be given.



A. B. Delfino and A. Buchs

The data submitted to the IG were the structure of 3 and its low-resolution mass spectrum, recorded at 70 eV. The input of structures and mass spectra is explained in our latest publication <sup>46)</sup>. The way structure 3 was numbered for input is shown below (4).



4

The numbering scheme can be arbitrary; different numbering schemes for the same structure will require different input cards, but will in no way affect the resultant structures and mechanisms proposed by the IG. Currently the program can accept up to fifty atoms, not including the hydrogen atoms.

The IG returns two kinds of output. On the main output, every step of the processes involved in the formation of an ionic species is shown in detail, with the structures of both the ionic and neutral (if any) species printed in linear notation. The linear notational scheme used for output is a non-canonical form of the Dendral notation of Lederberg <sup>39)</sup> with some minor modifications that we have mentioned in our latest paper <sup>46)</sup>.

The ionization range selected to ionize structure 3 extended from 800 to 1000. Thus, ionization took place only by removal of one of the lone-pair electrons from the carbonylic oxygen atom. A short-hand notation of the output for that process is:

ION 19    IMASS 270    INT 16%

Which means that ionization affected atom no. 19, yielding an ionic species with  $m/e$  270, appearing in the actual mass spectrum with a relative abundance of 16%.

From that molecular ion the program generated altogether 24 ions with different  $m/e$  values. Table 3 shows a list of the ions generated along with the number of mechanisms proposed for each ion. For demonstration purposes, the sizes of the allowed transition states were restricted to 4, 6 and 8-membered rings, and the ring closures to 4-membered rings. No restrictions were placed on the size of the radicals allowed to be transferred. A few examples will suffice to illustrate some of the proposed fragmentation mechanisms.

Table 3. Primary ions proposed by the IG program for 2-*n*-hexyl-methyl-*n*-decanoate

| <i>m/e</i> | Number of mechanisms proposed | Relative abundance in mass spectrum (%) <sup>a)</sup> |
|------------|-------------------------------|-------------------------------------------------------|
| 270        | 1                             | 16,0                                                  |
| 255        | 8                             | 0,4                                                   |
| 242        | 31                            | 0,6                                                   |
| 241        | 2                             | 2,5                                                   |
| 239        | 1                             | 4,8                                                   |
| 227        | 6                             | 2,7                                                   |
| 214        | 19                            | 1,1                                                   |
| 213        | 2                             | 5,9                                                   |
| 199        | 8                             | 15,0                                                  |
| 186        | 19                            | 68,0                                                  |
| 185        | 2                             | 7,7                                                   |
| 171        | 9                             | 19,0                                                  |
| 158        | 16                            | 100,0                                                 |
| 157        | 2                             | 14,0                                                  |
| 143        | 5                             | 30,0                                                  |
| 130        | 6                             | 2,5                                                   |
| 129        | 1                             | 17,0                                                  |
| 115        | 10                            | 28,0                                                  |
| 102        | 2                             | 2,6                                                   |
| 101        | 2                             | 22,0                                                  |
| 87         | 7                             | 99,0                                                  |
| 73         | 1                             | 2,9                                                   |
| 59         | 1                             | 7,9                                                   |
| 45         | 4                             | 2,9                                                   |

<sup>a)</sup> Mass spectrum no. 1898 from MSDC catalogue.

### 1. Simple $\alpha$ -Cleavages Yielding *m/e* 239 and *m/e* 59

The IG rationalizes the formation of these two ions in the following way:

ION 19 HOM 10,11 MKB 10,19 IMASS 239 NMASS 31 INT 4,8%

This reads: "Ionization on atom no. 19, homolysis between atom no. 10 and atom no. 11, bond formation between atom no. 10 and atom no. 19". The indications about the masses of the ionic (IMASS) and the neutral (NMASS) species, as well as about the relative abundance retrieved from the mass spectrum are presented in addition to the mechanism.

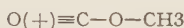
For *m/e* 59 the sequence of primitive operations forming the  $\alpha$ -cleavage mechanism is presented as follows:

ION 19 HOM 9,10 MKB 10,19 IMASS 59 NMASS 211 INT 7,9%

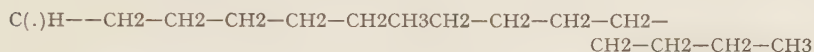
A. B. Delfino and A. Buchs

The structures of the ionic species and of the neutral species also appear in the output. The structures for the second mechanism described above are presented as shown below, indicating the site of the positive charge and of the radical.

*Ionic species*



*Neutral species*



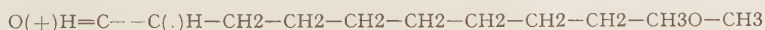
## 2. McLafferty Rearrangement Ions

Among the 19 mechanisms proposed for the formation of  $m/e$  186 and among the 16 mechanisms generated for  $m/e$  158, the two mechanisms corresponding to McLafferty rearrangements are present. For the rearrangement involving the  $n$ -hexyl group, the one which yields  $m/e$  186, the program did the following:

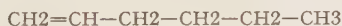
ION 19    6SMRTH-14    HOM 9,13    MKB 13,14  
IMASS 186    NMASS 84    INT 68%

The notation '6SMRTH-14' translates to: 'Short-lived 6-Membered Ring Transfer of a Hydrogen atom abstracted from atom no. 14 (to the existing radical site)'. The corresponding structures of ionic and neutral species are:

*Ionic species*



*Neutral species*



When the McLafferty rearrangement affects the  $n$ -octyl group, the ion with  $m/e$  158 is formed. The mechanism proposed by the ION GENERATOR is:

ION 19    6SMRTH-7    HOM 8,9    MKB 7,8  
IMASS 158    NMASS 112    INT 100%

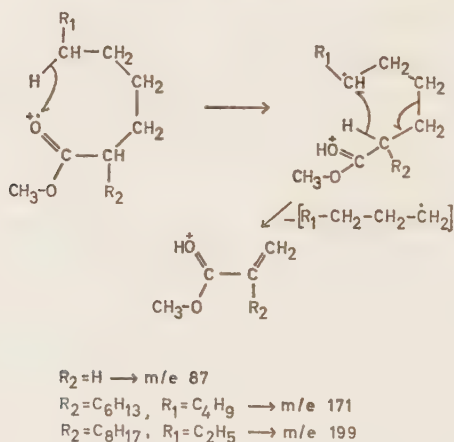
It is well known (47) that in the fragmentation of normal long-chain methyl esters the formation of fragments of the general formula  $(\text{CH}_2)_n\text{COOCH}_3^+$ , where  $n=2, 6, 10 \dots$ , is favored. Such ions are indicated at  $m/e$  87, 143, 199, 255 etc. in the mass spectra of unbranched methyl esters, *e. g.* spaced at intervals of 56 mass units. To explain their genesis, Djerassi

*et al.* <sup>48)</sup> propose mechanisms which involve two successive transfers of hydrogen atoms. With our example (3), among the mechanisms generated for  $m/e$  171 and for  $m/e$  199, one finds the two mechanisms which are analogous to the one proposed for the formation of  $m/e$  87 in the mass spectra of unbranched long-chain methyl esters.

$m/e$  171

ION 19 8SMRTH-5 6SMRTH-9 HOM 7,8  
IMASS 171 NMASS 99 INT 19%

This corresponds to the mechanism depicted in Scheme 1. Exactly the same

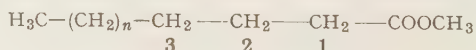


Scheme 1

mechanism is proposed for  $m/e$  199, with the difference that the transition state takes place in the  $n$ -hexyl radical. The mechanism is:

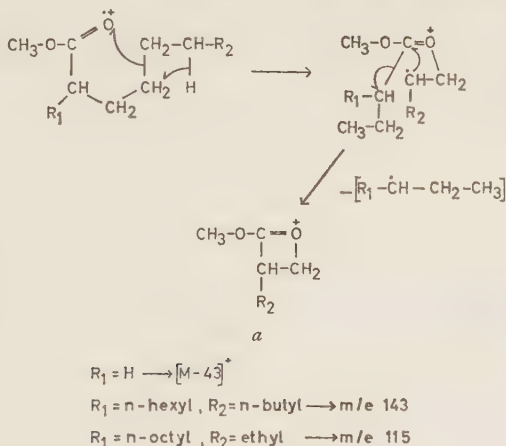
ION 19 8SMRTH-16 6SMRTH-9 HOM 13,14  
IMASS 199 NMASS 71 INT 15%

As an example of a mechanism proceeding through a 'long-lived' transition state, one can consider the expulsion of a  $n$ -propyl radical from long-chain methyl ester molecular ions. Djerassi *et al.* <sup>49)</sup> have postulated a mechanism which accounts for the observation that the three expelled carbon atoms are the closest neighbors of the carbonylic carbon atom, *i. e.* the carbon atoms numbered from 1 to 3 in the structure shown below.



A. B. Delfino and A. Buchs

Among the various mechanisms which the IG proposes for the occurrence of  $m/e$  143 and  $m/e$  115, *i. e.*  $(M - R - CH_2 - CH_3)^+$  where R is either a *n*-hexyl or a *n*-octyl radical, one finds exactly the same mechanism as that proposed by Djerassi. The detailed mechanism is depicted in Scheme 2. It yields  $m/e$  143 when the 6-membered 'long-lived' transition state operates



Scheme 2

on the *n*-hexyl radical connected to atom no. 9, and  $m/e$  115 in the case where the fragmentation starts with the same cyclic transition state, but on the octyl radical attached to carbon atom no. 9. For the fragmentation leading to  $m/e$  143, the output of the program can be summarized by the following shorthand notation:

|           |            |           |          |
|-----------|------------|-----------|----------|
| ION 19    | 6LMRTR85-7 | 4SMRTH-5  | HOM 9,10 |
| 4MRC 7,10 | IMASS 143  | NMASS 127 | INT 30%  |

This sequence of operations reads as follows:

'Ionization on atom no. 19, Longlived 6-Membered Ring Transfer of a Radical with mass 85 abstracted from atom no. 7, Short-lived 4-Membered Ring Transfer of a Hydrogen atom abstracted from atom no. 5, Homolysis between atom no. 9 and atom no. 10, 4-Membered Ring Closure between atom no. 7 and atom no. 10'.

The structure of the resulting ionic species (structure  $\alpha$  of Scheme 2) is printed in the following way:

*Ionic species*



As can be seen, one of the bonds in the ring is repeated twice; in this case it is the double bond, which is followed by the symbol 1 (the first non-hydrogen atom in the output string) when it is printed for the second time.

The mechanism leading to  $m/e$  115 is:

|            |             |           |          |
|------------|-------------|-----------|----------|
| ION 19     | 6LMRTR57-14 | 4SMRTH-16 | HOM 9,10 |
| 4MRC 16,10 | IMASS 115   | NMASS 155 | INT 28%  |

In this case the transferred radical is a *n*-butyl radical.

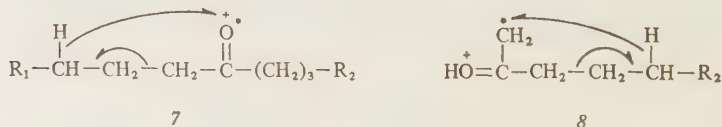
### C. Future Work with the Ion Generator Program

It is, of course, too early to predict all that might be accomplished with the IG program in the future. We have, however, some rather distinct preferences regarding what should be done which we would like to summarize here. First of all, the program, to be useful, must be able to work in the area of secondary ions, tertiary ions, etc., down to some lower limit on mass which would be set by the user. Hence, the IG should keep working on ions resulting from larger ions until this lower limit is reached.

Another area of interest lies in the comparison of sets of spectra of related compounds. Since the steps forming the various mechanism depend exclusively on the structures for their motivation, the types of primitive operations applied to a structure to yield mechanisms are sufficient for the purpose of comparing mechanisms from one structure to another. For example, simple  $\beta$ -cleavage (homolysis) followed by formation of a new bond (makebond), is the same mechanism in 5 as it is in 6.



Also, the McLafferty rearrangement (the sequence of operations: short-lived 6-membered transition state with transfer of a hydrogen atom, followed by  $\beta$ -cleavage) is the mechanism in 7 as well as in 8, a daughter ion of 7.



A. B. Delfino and A. Buchs

So the need is for a program that treats the output of the IG in such a way as to identify the common mechanisms which operate in a set of compounds to yield the important ions. This could lead to the generalization of certain electron book-keeping steps as ion-formation rationalizations.

Lastly, a program could perhaps be written to find the empirical constants needed to predict mass spectra. When a mass spectrometrists fragments a structure by the method of electron book-keeping, he has some rather definite ideas about the relative intensities that some of the ions so formed should exhibit. Writing a program to do this is by no means a short-range goal, but it is an enticing accomplishment toward which to work.

#### IV. References

- 1) Hites, R. A., Biemann, K.: *Anal. Chem.* **40**, 1217 (1968).
- 2) Sweeley, C. C., Ray, B. D., Wood, W. I., Holland, J. F., Krichewsky, M. I.: *Anal. Chem.* **42**, 1505 (1970).
- 3) Smith, D. H., Olsen, R. W., Walls, F. C., Burlingame, A. L.: *Anal. Chem.* **43**, 1796 (1971).
- 4) Henneberg, D., Schomberg, G.: In: *Advances in mass spectrometry* (ed. A. Quayle), Vol. 5, p. 605. London: Institute of Petroleum 1971.
- 5) Reynolds, W. E., Bacon, V. A., Bridges, J. C., Coburn, T. C., Halpern, B., Lederberg, J., Levinthal, E. C., Steed, E., Tucker, R. B.: *Anal. Chem.* **42**, 1122 (1970).
- 6) Venkataraghavan, R.: In: *Mass spectrometry. Techniques and applications* (ed. G. W. A. Milne), p. 1. New York: Wiley-Interscience 1971 (and references cited therein).
- 7) Desiderio, D. M.: In: *Mass spectrometry. Techniques and applications* (ed. G. W. A. Milne), p. 11. New York: Wiley-Interscience 1971 (and references cited therein).
- 8) Habfast, K.: In: *Advances in mass spectrometry* (ed. E. Kendrick), Vol. 4, p. 3. London: The Institute of Petroleum 1968.
- 9) Venkataraghavan, R., Board, R. D., Klimowsky, R., Amy, J. W., McLafferty, F. W.: In: *Advances in mass spectrometry*, Vol. 4, p. 65. London: The Institute of Petroleum 1968.
- 10) Wards, S. D.: In: *Mass spectrometry* (ed. D. H. Williams), p. 253. London: The Chemical Society 1971.
- 11) Burlingame, A. L.: In: *Recent developments in mass spectroscopy* (ed. K. Ogata and T. Hayakawa), p. 104; *Procs. Int. Conf. Kyoto, Japan, 8-12 Sept. 1969*. Baltimore-London-Tokyo: Univ. Park Press 1970.
- 12) Holmes, W. F., Holland, W. H., Parker, J. A.: *Anal. Chem.* **43**, 1806 (1971).
- 13) Hertz, H. S., Hites, R. A., Biemann, K.: *Anal. Chem.* **43**, 681 (1971) (and references cited therein).
- 14) Erni, F., Clerc, J. T.: *Helv. Chim. Acta* **55**, 489 (1972) (and references cited therein).
- 15) Jurs, P. C., Kowalski, B. R., Isenhour, T. L.: *Anal. Chem.* **41**, 21 (1969).
- 16) Jurs, P. C., Kowalski, B. R., Isenhour, T. L., Reilley, C. N.: *Anal. Chem.* **41**, 690 (1969).
- 17) Jurs, P. C., Kowalski, B. R., Isenhour, T. L.: *Anal. Chem.* **41**, 695 (1969).
- 18) Kowalski, B. R., Jurs, P. C., Isenhour, T. L., Reilley, C. N.: *Anal. Chem.* **41**, 1945 (1969).

- 19) Jurs, P. C., Kowalski, B. R., Isenhour, T. L., Reilley, C. N.: *Anal. Chem.* **41**, 1949 (1969).
- 20) Jurs, P. C.: *Anal. Chem.* **42**, 1633 (1970).
- 21) Jurs, P. C.: *Anal. Chem.* **43**, 22 (1971).
- 22) Jurs, P. C.: *Appl. Spectry.* **25**, 483 (1971).
- 23) Jurs, P. C.: *Anal. Chem.* **43**, 1812 (1971).
- 24) Petterson, B., Ryhage, R.: *Anal. Chem.* **39**, 790 (1967).
- 25) Petterson, B., Ryhage, R.: *Arkiv Kemi* **26**, 293 (1967).
- 26) Biemann, K., McMurray, W., Fennessey, P. V.: *Tetrahedron Letters* **1966**, 3997.
- 27) Biemann, K., Cone, C., Webster, B. R., Arsenault, G. P.: *J. Am. Chem. Soc.* **88**, 5598 (1966).
- 28) Senn, M., Venkataraghavan, R., McLafferty, F. W.: *J. Am. Chem. Soc.* **88**, 5592 (1966).
- 29) Barber, M., Powers, P., Wallington, M. J., Wolstenholme, W. A.: *Nature* **212**, 784 (1966).
- 30) Biemann, K., Fennessey, P. V.: *Chimia (Aarau)* **21**, 266 (1967).
- 31) Fennessey, P. V.: Ph. D. thesis, Massachusetts Institute of Technology. Cambridge, Mass. 1968.
- 32) Crawford, L. R., Morrison, J. D.: *Anal. Chem.* **43**, 1790 (1971).
- 33) Crawford, L. R., Morrison, J. D.: *Anal. Chem.* **41**, 994 (1969).
- 34) Crawford, L. R., Morrison, J. D.: *Anal. Chem.* **40**, 1469 (1968).
- 35) McLafferty, Mass spectral correlations. Washington, D. C.: American Chemical Society 1963.
- 36) Smith, D. H.: *Anal. Chem.* **44**, 536 (1972).
- 37) Venkataraghavan, R., McLafferty, F. W., Van Lear, G. E.: *Org. Mass Spectrom.* **2**, 1 (1969).
- 38) Riemann, K., McMurray, W.: *Tetrahedron Letters* **1965**, 647.
- 39) Lederberg, J.: In: *Topology of molecules*, p. 37; The mathematical sciences. Cambridge, Mass.—London: MIT Press, 1969.
- 40) Lederberg, J., Sutherland, G. L., Buchanan, B. G., Feigenbaum, E. A., Robertson, A. V., Duffield, A. M., Djerassi, C.: *J. Am. Chem. Soc.* **91**, 2973 (1969).
- 41) Duffield, A. M., Robertson, A. V., Djerassi, C., Buchanan, B. G., Sutherland, G. L., Feigenbaum, E. A., Lederberg, J.: *J. Am. Chem. Soc.* **91**, 2977 (1969).
- 42) Schroll, G., Duffield, A. M., Djerassi, C., Buchanan, B. G., Sutherland, G. L., Feigenbaum, E. A., Lederberg, J.: *J. Am. Chem. Soc.* **91**, 7440 (1969).
- 43) Buchs, A., Duffield, A. M., Schroll, G., Djerassi, C., Delfino, A. B., Buchanan, B. G., Sutherland, G. L., Feigenbaum, E. A., Lederberg, J.: *J. Am. Chem. Soc.* **92**, 6831 (1970).
- 44) Buchs, A., Delfino, A. B., Duffield, A. M., Djerassi, C., Buchanan, B. G., Feigenbaum, Lederberg, J.: *Helv. Chim. Acta* **53**, 1394 (1970).
- 45) Buchanan, B. G., Feigenbaum, E. A., Lederberg, J.: Stanford Artificial Intelligence Project, Memo AIM-145, Report No. CS-221 (1971).
- 46) Delfino, A. B., Buchs, A.: *Helv. Chim. Acta* **55**, 2017 (1972).
- 47) Spiteller, G., Spiteller-Friedmann, M., Houriet, R.: *Monatsh. Chem.* **97**, 121 (1966).
- 48) Budzikiewicz, H., Djerassi, C., Williams, D. H.: In: *Mass spectrometry of organic compounds*, p. 179. San-Francisco: Holden-Day, Inc. 1967.
- 49) Budzikiewicz, H., Djerassi, C., Williams, D. H.: In: *Mass spectrometry of organic compounds*, p. 180. San-Francisco: Holden-Day, Inc. 1967.

Appendix B.

This article appeared in the Helvetica Chimica Acta during 1972. Therein, the charge localization concept in conjunction with mechanism generation was presented.

# HELVETICA CHIMICA ACTA

---

EDENDA CURAT SOCIETAS CHIMICA HELVETICA

Volumen 55  
Fasciculus sextus  
1972

194. Heuristic Programming as an Ion Generator  
in Mass Spectrometry

I. Generation of Primary Ions with Charge Localization

by Allan B. Delfino and Armand Buchs

Department of Physical Chemistry,  
University of Geneva, 1211 Geneva 4, Switzerland.

(3 V 72)

VERLAG HELVETICA CHIMICA ACTA  
4002 BASEL, SCHWEIZ

---

SEPARATUM

**194. Heuristic Programming as an Ion Generator  
in Mass Spectrometry**  
**I. Generation of Primary Ions with Charge Localization**

by **Allan B. Delfino** and **Armand Buchs**

Department of Physical Chemistry,  
University of Geneva, 1211 Geneva 4, Switzerland.

(3 V 72)

*Résumé.* Un programme connu sous le nom de *ION GENERATOR* a été élaboré. Pour le moment ce programme est capable de créer, à partir de n'importe quelle molécule organique, les ions primaires résultant de la fragmentation de l'ion moléculaire et de proposer des mécanismes de fragmentation pour expliquer la formation des ions.

In this paper we present a heuristic program we have devised to simulate the formation of ions in the ion source of a mass spectrometer. Our program, the *ION GENERATOR*, acts in much the same way as the chemist who is trying to rationalize ion formations with the help of paper and pencil; it is based on the well known method

of electron book-keeping. The program is expandable to include in the future a great proportion of the mechanistic steps delineated for example in the textbook by *Djerassi et al.* [1]. With such a program one can expect to generate, from any molecule, a number of important ions present in a mass spectrum. The advantages of using a computer to look for fragmentation mechanisms are manyfold, but we would like to name two which we consider to be of primary importance. Since the electron book-keeping method is tedious, the chemist usually stops when he has realized a plausible rationalization of the fragmentation. But the computer will realize an *exhaustive* and *irredundant* set of plausible rationalizations that are consonant with the structures of the formed ions and that are not chemically unsound. Hence, the chemist who utilises the program could then design, on the basis of the proposed fragmentation paths, labelling experiments to determine which of the sets present indeed the best rationalizations, or which fragmentation modes are occurring simultaneously.

The second reason for interest in this kind of research is the fact that the methods of electron book-keeping were handed down by examples, after presenting of a few rules. If one wants a computer to handle electron book-keeping, a much more rigorous presentation must be worked out. One is compelled to generalize, as much as can be done, each and every step of a fragmentation process. Besides the research described here, another area of research where heuristic search techniques are applied to mass spectrometry has recently been published by *Buchanan et al.* [2].

In order to generate molecular ions from a molecule and to fragment them to form all plausible primary ions, the *ION GENERATOR* program must know the motivations and methods of altering structures, which we call primitive operations. By applying these primitive operations in all plausible sequences to a molecule, or to an ion, until the program finds a new ionic species different in mass from the parent species, it builds sets of primitive operations which represent fragmentation mechanisms and which yield ions of various masses.

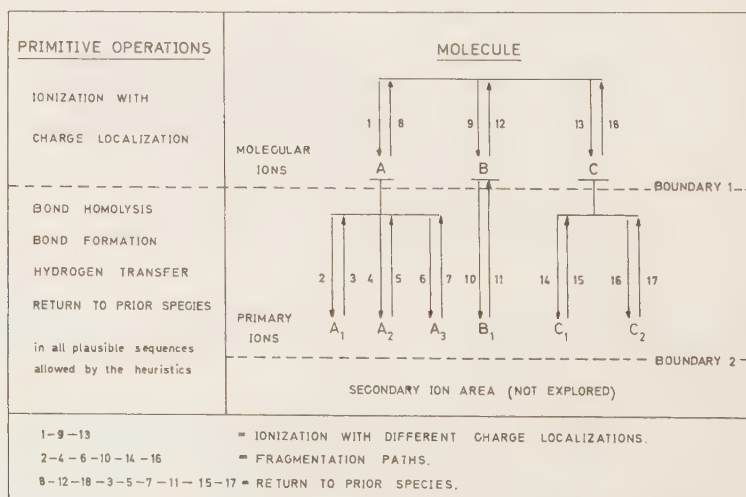


Figure 1. Sequence of operations to ionize and fragment molecules.

The way the program applies the primitive operations will be illustrated with an example, the ionization and fragmentation of 5-diethylamino-pentan-2-one (**1**).

The input of the structure is explained by using the example 5-diethylamino-pentan-2-one (**1**). The atoms are first numbered in an *arbitrary* way. To make the input coding as short as possible, the longest chain of singly bonded atoms should be searched for and the atoms of that chain should be numbered in sequence. A convenient way to number the atoms of structure **1** is shown below (**2**).

1) The program can generate any depth in the ionic area, but lack of heuristics beyond boundary 2 (Fig. 1) prevents excursion in the ionic area retaining plausibility.

of a *BOND CARD* is used to differentiate between connections through a group of atoms and connections between two atoms only; the digit '1' is punched in that column for a group and the digit '0' for connections between two atoms only. The next two sets of three columns are used for the numbers assigned to the first and to the last atom of the chain. The eight and last column transmits the information about the bond order, i.e. 1, 2 or 3 for single, double and triple bonds, respectively.

With the numbering shown in **2** the *BOND CARDS* to input the structure of 5-diethylamino-pentan-2-one are the following (**4**):

| <i>BOND CARD No</i> | <i>Columns</i> |   |   |   |   |   |   |   |
|---------------------|----------------|---|---|---|---|---|---|---|
|                     | 1              | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1                   | 1              |   |   | 1 |   |   | 8 | 1 |
| 2                   | 0              |   |   | 9 |   |   | 3 | 1 |
| 3                   | 0              |   |   | 9 |   | 1 | 0 | 1 |
| 4                   | 0              |   |   | 7 |   | 1 | 1 | 2 |

4

These four *BOND CARDS* transmit the following information:

*CARD No. 1:* From atom No. 1 to atom No. 8 there is a group of atoms where each atom is connected to the following one in number by a single bond.

*CARD No. 2:* Atom No. 9 is connected to atom No. 3 by a single bond.

*CARD No. 3:* Atom No. 9 is connected to atom No. 10 by a single bond.

*CARD No. 4:* Atom No. 7 is connected to atom No. 11 by a double bond.

These cards can be given to the program in any sequence. The number of *BOND CARDS* needed to input a structure depends on how the atoms are numbered. The number of cards can always be minimized by choosing the appropriate numbering sequence.

*Notational scheme for the output of chemical structures.* - To present the structures of the ions and of the neutral species formed as a result of applying the primitive operations to a molecule or to an ion, a notation similar to *Lederberg's DENDRAL* notation is used [3]. The differences between the linear notation used here and the *DENDRAL* notation are that our notation shows the hydrogen atoms of the structure and also indicates on which atoms radical sites and positive charges are found. Moreover, the notation used here has no canonical order. The *DENDRAL* notation has recently been nicely explained by *Buchanan et al.* in a review paper [4].

The two general rules of the notation used by the *ION GENERATOR* are:

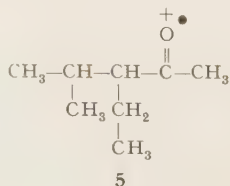
*Rule 1:* An atom is printed along with its hydrogen atoms, followed by its free bonds.

*Rule 2:* The radical(s) attached to the free bond(s) of an atom are printed in the same order as were the bond(s) according to rule 1.

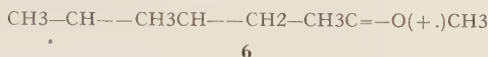
The following examples illustrate both rules.



As an example of a ionic species one can consider the following structure (**5**):



The linear notation for this structure can be written as (6).



*Ionization and charge localization.* - In mass spectrometry, rationalization of bond cleavages by the electron book-keeping method assumes that in a molecular ion the charge is preferentially located at favorable sites, just prior to fragmentation. It is not unreasonable to think about the primary ions as originating from different molecular ions. These molecular ions could differ with respect to the charge locus, to their relative stability, linked to their tendency to fragment before they can rearrange. Functional groups containing heteroatoms generally provide the best loci for the positive charge. Carbon-carbon multiple bonds also play an important role while carbon-carbon single bonds and carbon-hydrogen bonds are much less effective. The program has been designed to know that the most likely occurring molecular ions are those where the active site arises by removal of a non-bonding electron, followed by those in which the expelled electron is a  $\pi$  type electron. It also knows which heteroatoms are more effective than others to stabilize the positive charge. For each kind of electron type and for each environment a unique ionization threshold value has been semi-empirically chosen. Table 1 shows threshold values for a number of struc-

Table 1. *Threshold values for ionization.*

| Substructure               | -N<  | -NH | -NH <sub>2</sub> | >C-S | >C=O | -S- | -O- | -SH |
|----------------------------|------|-----|------------------|------|------|-----|-----|-----|
| Ionization threshold value | 1000 | 950 | 900              | 850  | 800  | 750 | 700 | 650 |

| Substructure               | -OH | -C=C- | >C=C< | $\begin{array}{c}   \\ -\text{C}- \\   \end{array}$ | $\begin{array}{c}   \\ -\text{CH}- \\   \end{array}$ | -CH <sub>2</sub> - | $\begin{array}{c}   \\ \text{H}-\text{C}-\text{H} \\   \\ \text{H} \end{array}$ |
|----------------------------|-----|-------|-------|-----------------------------------------------------|------------------------------------------------------|--------------------|---------------------------------------------------------------------------------|
| Ionization threshold value | 600 | 550   | 500   | 450                                                 | 400                                                  | 350                | 300                                                                             |

tural environments. These values, which range arbitrarily from 300 to 1000, allow the user of the program to select, by means of an option card, the kind(s) of ionization that he wants the structure to undergo. For example, if one wants ionization to take place only on the heteroatoms present in a molecule, one will select a range extending from 600 to 1000. In a difunctional molecule, it is possible to generate the ions arising from charge localization on the two heteroatoms separately. The program would have to be run twice, each time with the appropriate range for the ionization threshold values. For any chosen range, including the whole range, i.e. 300 to 1000, the *ION GENERATOR* builds the molecular ions and fragments them in order of decreasing threshold values.

*Bond homolysis and bond formation.* - The only motivation currently known to the program to homolyze a particular bond is the availability of a radical site in a  $\beta$  position. This allows, afterwards, the formation of a new bond either in the ionic or in the neutral species. Another requirement a particular bond must satisfy in order to be fissionable is that it must not have been formed by the *ION GENERATOR*

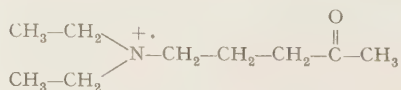
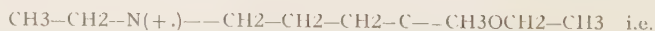
in a previous operation of the ongoing fragmentation mode. Regarding C-H bond cleavages, the program considers all such bonds as equivalent when they originate from the same atom. Only one of the equivalent bonds will be homolyzed. Each time an ion has been formed, the program searches the mass spectrum for the relative abundance of the peak corresponding to the ion and prints it out, along with the structures of the ionic and neutral (if any) species. Even homolytic cleavages which do not change the mass of an ionic species are performed by the program. This affords a way to create a new radical site which can then play the trigger role for the next step of the fragmentation mode. However, homolysis performed on a double bond is not allowed to be the last step of a sequence of primitive operations; a fragmentation mechanism always ends with an ionic species having a mass which is smaller than the mass of the precursor ion.

*Transfer of a hydrogen atom to a radical site via cyclic transition states of various sizes.* - The tendency of a radical site to form a new bond can also be satisfied by the transfer of a hydrogen atom to the radical site. This creates a new radical site which is then able either to induce  $\beta$ -cleavage or to trigger a new hydrogen transfer. The sizes of the cyclic transition states which are presently allowed are from four- to eight-membered rings. Some loose restrictions are placed on the sizes which are allowed. Thus, for example, a hydrogen atom cannot be abstracted from a carbon atom bearing a double bond, or from an atom on which a hydrogen atom has already been transferred during a previous operation of the ongoing mechanism. Situations leading to blatant strain are also avoided. Moreover, the *ION GENERATOR* allows only two transition states to occur in any sequence of primitive operations. This restriction is based on the fact that extensive rearrangement before fragmentation, leading to drastic scrambling of the hydrogen atoms, is contrary to the ability of mass spectrometry to elucidate structures.

*Mechanisms proposed by the ION GENERATOR for the formation of primary ions from 5-diethylaminopentan-2-one (1).* - The ionization range which was selected extended from 750 to 1000. From Table 1 it can be seen that ionization then occurs only on the two heteroatoms. Two molecular ions were thus generated, one by lone-pair ionization on the nitrogen atom (atom No. 3 according to the numbering shown in 2) and the other by lone-pair ionization on the oxygen atom (atom No. 11).

The program returns two kinds of outputs. On the main output every step of the various processes leading to the formation of an ion is shown in detail, with the structures of both the ionic and neutral (if any) species printed in the linear notation. A second output shows what occurred in a shorthand notation. The structures of the two molecular ions as they are printed are shown below with their respective translations (7 and 8).

*Molecular ion No. 1*

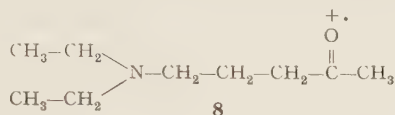
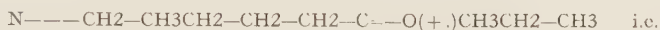


The abbreviated notation for that process is printed as follows:

ION 3 IMASS 157 INT 5,1% NMASS 0, which means:

'ionization affects atom No. 3, mass of the ionic species is 157, relative abundance of the peak found in the actual mass spectrum is 5,1%, mass of the neutral fragment lost is zero'.

*Molecular ion No. 2*



The abbreviated notation is:

ION 11 IMASS 157 INT 5,1% NMASS 0

From these two molecular ions the program generated altogether 16 ions of different masses by various mechanisms. The list of the ions which were generated is

Table 2. *Number of mechanisms proposed for the formation of primary ions from 5-diethylaminopentan-2-one.*

| Ionization on nitrogen |   |   |             |                            | Ionization on oxygen |   |   |                            |                           |
|------------------------|---|---|-------------|----------------------------|----------------------|---|---|----------------------------|---------------------------|
| Mechanisms involving   |   |   |             | Total number of mechanisms | Mechanisms involving |   |   | Total number of mechanisms | Intensity <sup>a)</sup> % |
| 0                      | 1 | 2 | 0           |                            | 1                    | 2 |   |                            |                           |
| H transfers            |   |   | H transfers |                            |                      |   |   |                            |                           |
| <i>m/e</i>             |   |   |             |                            |                      |   |   |                            |                           |
| 43                     |   |   |             |                            | 1                    | 0 | 2 | 3                          | 5,2                       |
| 44                     |   |   |             |                            | 0                    | 0 | 1 | 1                          | 5,1                       |
| 58                     |   |   |             |                            | 0                    | 1 | 1 | 2                          | 26,4                      |
| 71                     |   |   |             |                            | 0                    | 0 | 2 | 2                          | 3,6                       |
| 72                     | 0 | 0 | 3           | 3                          |                      |   |   |                            | 3,4                       |
| 73                     | 0 | 1 | 1           | 2                          |                      |   |   |                            | 1,0                       |
| 85                     |   |   |             |                            | 0                    | 0 | 1 | 1                          | 30,5                      |
| 86                     | 1 | 0 | 0           | 1                          | 0                    | 1 | 2 | 3                          | 100,0                     |
| 87                     | 0 | 1 | 1           | 2                          |                      |   |   |                            | 11,0                      |
| 100                    | 0 | 0 | 3           | 3                          |                      |   |   |                            | 0,8                       |
| 114                    | 0 | 1 | 1           | 2                          |                      |   |   |                            | 0,6                       |
| 115                    | 0 | 1 | 0           | 1                          |                      |   |   |                            | 0,1                       |
| 128                    | 0 | 0 | 3           | 3                          | 0                    | 0 | 1 | 1                          | 0,2                       |
| 129                    | 0 | 1 | 0           | 1                          |                      |   |   |                            | 0,1                       |
| 142                    | 1 | 0 | 0           | 1                          | 1                    | 0 | 0 | 1                          | 3,6                       |
| 156                    | 2 | 0 | 0           | 2                          |                      |   |   |                            | 0,8                       |
| 157                    |   |   |             | 1                          |                      |   |   | 1                          | 5,1                       |
|                        |   |   |             | 22                         |                      |   |   | 15                         |                           |

<sup>a)</sup> Intensity values refer to relative abundances with intensity of the base peak = 100%.

shown in Table 2 along with the number of non-equivalent mechanisms proposed for the formation of each ion. In the structure of **1**, carbon atoms No. 1 and No. 2 are respectively equivalent to carbon atoms No. 10 and No. 9. Equivalent mechanisms involving these carbon atoms have been counted once only. Moreover, when transfer of hydrogen atoms occurs, only those mechanisms which could be, if desired, subjected to direct experimental proof by means of labelling with deuterium, have been retained in Table 2. Mechanisms in which all the rearrangements of hydrogen atoms are internal to the product ion, or to the expelled neutral fragment, have been rejected manually.

Table 3. *Mechanisms proposed for the fragmentation of 5-diethylamino-pentan-2-one down to primary ions*

| <i>m/e</i>        | Mechanisms proposed |          |          |         |          | Order |
|-------------------|---------------------|----------|----------|---------|----------|-------|
| 43                | ION 11              | HOM 6,7  | MKB 7,11 |         |          | 25    |
|                   | ION 11              | 6MRTH4   | 6MRTH8   | HOM 6,7 | MKB 7,11 | 29    |
|                   | ION 11              | 8MRTH9   | 8MRTH8   | HOM 6,7 | MKB 7,8  | 37    |
| 44                | ION 11              | 8MRTH9   | 5MRTH5   | HOM 6,7 |          | 34    |
| 58                | ION 11              | 6MRTH4   | HOM 5,6  |         |          | 26    |
|                   | ION 11              | 8MRTH9   | 4MRTH4   | HOM 5,6 |          | 31    |
| 71                | ION 11              | 6MRTH4   | 4MRTH6   | HOM 4,5 | MKB 5,6  | 27    |
|                   | ION 11              | 8MRTH9   | 6MRTH6   | HOM 4,5 | MKB 5,6  | 36    |
| 72                | ION 3               | 4MRTH5   | 5MRTH9   | HOM 3,4 | MKB 3,9  | 15    |
|                   | ION 3               | 5MRTH6   | 6MRTH9   | HOM 3,4 | MKB 3,9  | 19    |
|                   | ION 3               | 7MRTH8   | 8MRTH9   | HOM 3,4 | MKB 3,9  | 22    |
| 73                | ION 3               | 4MRTH10  | 6MRTH5   | HOM 3,4 |          | 10    |
|                   | ION 3               | 4MRTH5   | HOM 3,4  |         |          | 13    |
| 85                | ION 11              | 8MRTH9   | 5MRTH5   | HOM 3,4 | MKB 5,4  | 35    |
| 86                | ION 3               | HOM 4,5  | MKB 3,4  |         |          | 4     |
|                   | ION 11              | 6MRTH4   | 4MRTH9   | HOM 3,4 |          | 28    |
|                   | ION 11              | 8MRTH9   | HOM 3,4  |         |          | 30    |
|                   | ION 11              | 8MRTH9   | 4MRTH2   | HOM 3,4 |          | 33    |
| 87                | ION 3               | 5MRTH6   | HOM 4,5  |         |          | 17    |
|                   | ION 3               | 4MRTH10  | 7MRTH6   | HOM 4,5 |          | 12    |
| 100               | ION 3               | 4MRTH10  | 5MRTH4   | HOM 5,6 | MKB 4,5  | 8     |
|                   | ION 3               | 5MRTH6   | 4MRTH4   | HOM 5,6 | MKB 4,5  | 18    |
|                   | ION 3               | 7MRTH8   | 6MRTH4   | HOM 5,6 | MKB 4,5  | 21    |
| 114               | ION 3               | 4MRTH5   | HOM 6,7  | MKB 5,6 |          | 14    |
|                   | ION 3               | 4MRTH10  | 6MRTH5   | HOM 6,7 | MKB 5,6  | 11    |
| 115               | ION 3               | 7MRTH8   | HOM 6,7  |         |          | 20    |
| 128               | ION 3               | 4MRTH10  | 5MRTH4   | HOM 3,9 | MKB 3,4  | 7     |
|                   | ION 3               | 4MRTH10  | 5MRTH2   | HOM 3,9 | MKB 2,3  | 9     |
|                   | ION 3               | 4MRTH1   | 5MRTH4   | HOM 2,3 | MKB 3,4  | 16    |
|                   | ION 11              | 8MRTH9   | 4MRTH4   | HOM 3,9 | MKB 3,4  | 32    |
| 129               | ION 3               | 4MRTH10  | HOM 3,9  |         |          | 6     |
| 142               | ION 3               | HOM 1,2  | MKB 2,3  |         |          | 2     |
|                   | ION 11              | HOM 7,8  | MKB 7,11 |         |          | 24    |
| 156               | ION 3               | HOM 2,2' | MKB 2,3  |         |          | 3     |
|                   | ION 3               | HOM 4,4' | MKB 3,4  |         |          | 5     |
| 157               | ION 3               |          |          |         |          | 1     |
| (M <sup>+</sup> ) | ION 11              |          |          |         |          | 23    |

As can be seen from Table 2, the program proposed 35 fragmentation mechanisms for the formation of the 16 aforementioned ions. Some of the ions, those with  $m/e$  43, 44, 58, 71 and 85, originate, according to the *ION GENERATOR*, only from the molecular ion No. 2. Others, with  $m/e$  72, 73, 87, 100, 114, 115, 129 and 156, have been found only through mechanisms starting with the molecular ion No. 1. Finally, those ions with  $m/e$  86, 128 and 142 have been explained by the program starting with both molecular ions. Table 2 also shows, for each ion, how many mechanisms involved no hydrogen transfer before cleavage, those which involved the transfer of one hydrogen atom and those which involved the transfer of two hydrogen atoms before bond homolysis occurred to yield the ion.

All of the 35 proposed mechanisms are shown, in shorthand notation, in Table 3 along with the order in which the program presented them. For every mechanism listed in Table 3, a labelling experiment could be designed to substantiate or reject the proposed hydrogen transfer<sup>2)</sup>.

Some of the mechanisms proposed by the *ION GENERATOR* will now be illustrated with examples.

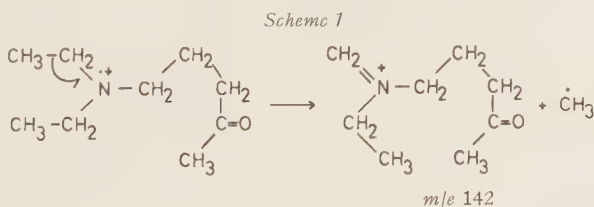
*α-cleavage next to the nitrogen atom.* -

a)  $m/e$  142

ION 3 HOM 1,2 MKB 2,3 IMASS 142 INT 3,6% NMASS 15

i.e.: 'ionization on atom No. 3, homolysis between atom No. 1 and atom No. 2, bond formation between atom No. 2 and atom No. 3'. The information concerning the masses of the ionic and neutral species, and the intensity of the corresponding peak found in the mass spectrum are self explanatory.

The shorthand notation shown above corresponds to the mechanism outlined in Scheme 1. The detailed output shows the following structures:



IONIC SPECIES:  $N(+) = -CH_2CH_2-CH_2-CH_2-C- = CH_3OCH_2-CH_3$

NEUTRAL SPECIES:  $C(•)H_3$

An equivalent mechanism is proposed for the formation of  $m/e$  142 by  $\alpha$ -cleavage. It involves loss of carbon atom No. 10 instead of carbon atom No. 1. As was mentioned before, whenever two mechanisms proposed by the program are equivalent, only one of them is retained in the list given in Table 3.

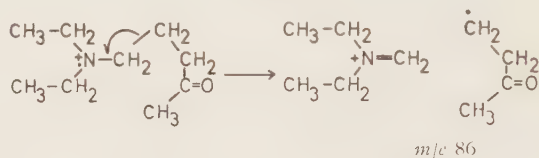
b)  $m/e$  86

ION 3 HOM 4,5 MKB 3,4 IMASS 86 INT 100% NMASS 71

<sup>2)</sup> For some of the mechanisms with double hydrogen transfer, labelling experiments could only verify one of the hydrogen transfers.

This mechanism is shown in Scheme 2. The following structures are shown in the output:

Scheme 2

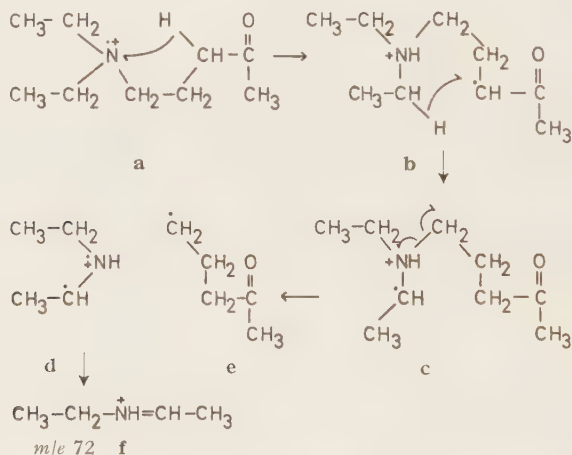


IONIC SPECIES:  $N(+)- = -CH_2-CH_3CH_2CH_2-CH_3$

NEUTRAL SPECIES:  $C(.)H_2-CH_2-C = -OCH_3$

*Cleavage of molecular ion No. 1 with transfer of hydrogen atoms.* - An example in which two consecutive transfers of hydrogen atoms to a radical site occur before a

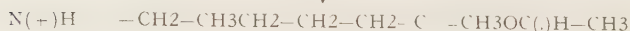
Scheme 3



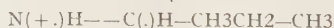
a ↓



b ↓

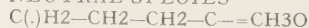


c ↓



d

NEUTRAL SPECIES



e

IONIC SPECIES



f

neutral fragment is expelled is afforded by the mechanism proposed for the formation of  $m/e$  72.

ION 3 5MRTH6 6MRTH9 HOM 3,4 MKB 3,9 IMASS 72 INT 3,4% NMASS 85  
i.e. 'ionization on atom No. 3, transfer of a hydrogen atom from atom No. 6 *via* a five-membered transition state, transfer of a hydrogen atom from atom No. 9 *via* a six-membered transition state, bond homolysis between atom No. 3 and atom No. 4, bond formation between atom No. 3 and atom No. 9'.

The corresponding conventional way of presenting that sequence of operations is illustrated in Scheme 3. For that sequence the structures depicted in Scheme 3 (a to f) are printed by the program as follows.

*Fragmentation starting with molecular ion No. 2.* -- The two  $\alpha$ -cleavages next to the carbonyl function are the only mechanisms without hydrogen transfer proposed by the ION GENERATOR for the fragmentation of molecular ion No. 2.

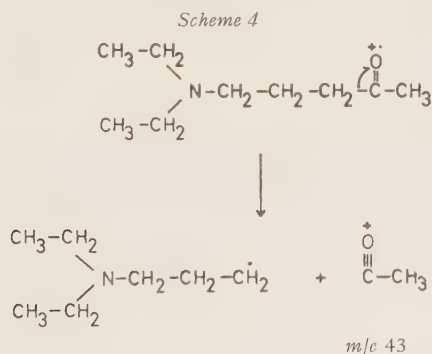
a)  $m/e$  43

ION 11 HOM 6,7 MKB 7,11 IMASS 43 INT 52,8% NMASS 114

IONIC SPECIES:  $C = -O(+)CH_3$

NEUTRAL SPECIES:  $C(.)H_2-CH_2-CH_2-N--CH_2-CH_3CH_2-CH_3$

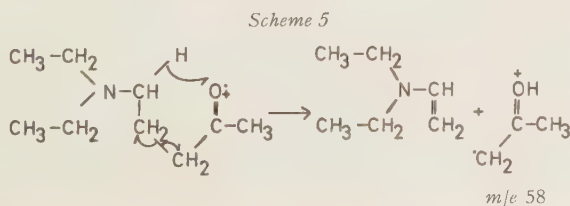
This corresponds to the mechanism depicted in Scheme 4.



Cleavage on the other side of the carbonyl function was also proposed to yield  $m/e$  142. The abbreviated notation for that fragmentation is:

ION 11 HOM 7,8 MKB 7,11 IMASS 142 INT 3,6% NMASS 15

b)  $m/e$  58



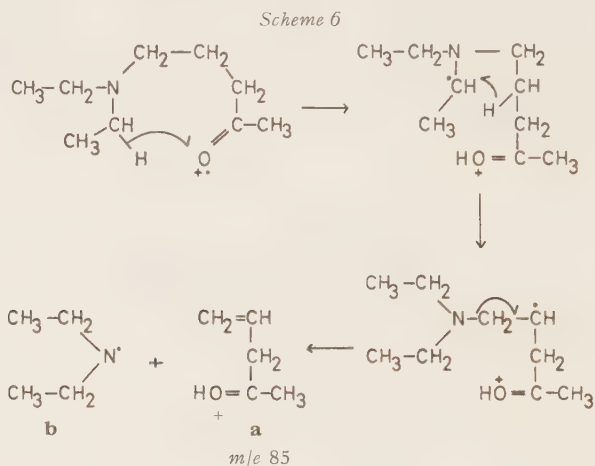
The *McLafferty* rearrangement shown in Scheme 5, which leads to *m/e* 58, was proposed in the following way:

ION 11 6MRTH4 HOM 5,6 IMASS 58 INT 26,4% NMASS 99

IONIC SPECIES:  $O(+)\text{H} = \text{C}--\text{CH}_3\text{C}(.)\text{H}_2$

NEUTRAL SPECIES:  $\text{CH}_2 = \text{CH}-\text{N}--\text{CH}_2-\text{CH}_3\text{CH}_2-\text{CH}_3$

As can be seen from Table 3, the program proposed only one mechanism for the formation of *m/e* 85. The fragmentation was visualized as starting from the molecular



ion No. 2 to yield ion **a** shown in Scheme 6. The abbreviated notation corresponding to the mechanism illustrated in Scheme 6 and the structures of the ionic and neutral species are:

ION 11 8MRTH9 5MRTH5 HOM 3,4 MKB 4,5 IMASS 85 INT 30,7% NMASS 72

IONIC SPECIES:  $O(+)\text{H} = \text{C}--\text{CH}_3\text{CH}_2-\text{CH} = \text{CH}_2$   
(structure **a** in Scheme 6)

NEUTRAL SPECIES:  $\text{N}(.)--\text{CH}_2-\text{CH}_3\text{CH}_2-\text{CH}_3$   
(structure **b** in Scheme 6)

A high-resolution measurement showed the signal at *m/e* 85 to be actually a singlet, due to an ion with the elemental composition  $\text{C}_5\text{H}_9\text{O}$ .

Table 3 shows also the order in which the ions were generated. The first ion to be created was, as has already been mentioned, the molecular ion No. 1. After that, the program generated all the ions arising from that form of the molecular ion, before building the other form of the molecular ion. Mechanisms which yield the same ion, from the same molecular ion, do not necessarily follow each other as is apparent from Table 3. This is due to the several points of view taken by the program in order to generate *all* plausible mechanisms.

By comparing the actual mass spectrum (Figure 2) of 5-diethylamino-pentan-2-one with the list of the ions shown in Table 2, it can be concluded that most of the important ions were generated. Although 22 of the 35 proposed mechanisms involve a double hydrogen transfer, they are plausible mechanisms and cannot be rejected *a priori*, without disproving them by means of deuterium labelling experiments. We are convinced that the *ION GENERATOR* has the potentiality to discover new mechanisms which have been missed in past studies of mass spectral fragmentation. The next step will be to implement more sophisticated primitive operations and to allow the program to explore any depth in the ionic area.

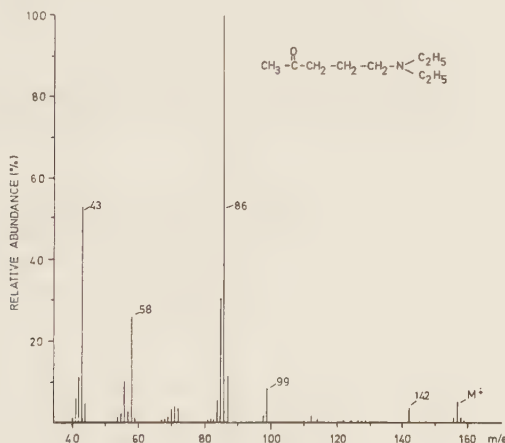


Figure 2: Mass spectrum of 5-diethylamino-pentan-2-one (70 eV).

### Experimental Part

The computer program was developed for the CDC 3800 computer at the Computer Center of the University of Geneva. It is written in the FORTRAN programming language. To fragment the structure of 5-diethylaminopentan-2-one down to primary ions the computer needed approximately 3 minutes. The low-resolution mass spectrum of **1** was recorded at 70 eV with a Varian MAT CH-4 mass spectrometer. High-resolution measurements were performed with a Varian MAT SM-1 instrument.

Support of part of this work by a grant (No 2.281.70) of the *Fonds National Suisse de la Recherche Scientifique* is gratefully acknowledged. We also thank Prof. B. Levrat for financial assistance to one of us (A.D.) and Mrs. F. Kloeti for her technical assistance with the mass spectrometer and for the drawing of the figures.

### BIBLIOGRAPHY

- [1] H. Budzikiewicz, C. Djerassi & D. H. Williams, 'Mass Spectrometry of Organic Compounds', Holden-Day, San Francisco 1967.
- [2] B. G. Buchanan, E. A. Feigenbaum & J. Lederberg, MEMO AIM-145, Report No. CS-221, Stanford Artificial Intelligence Project, 1971.
- [3] J. Lederberg, 'Topology of Molecules', in The Mathematical Sciences', M.I.T. Press, Cambridge, Mass. 1969, pp. 37-51.
- [4] B. G. Buchanan, A. M. Duffield & A. V. Robertson, in 'Mass Spectrometry. Techniques and Applications'. Ed. G.W.A. Milne, Wiley-Interscience, New-York 1971, pp. 121-178.

Appendix C.

This article appeared in Organic Mass Spectrometry during this year, 1974. In it is found a discussion of the ION GENERATOR operating on cyclic molecules, but restricted to the generation of primary ions.

## HEURISTIC PROGRAMMING AS AN ION GENERATOR IN MASS SPECTROMETRY

### II\*—GENERATION OF PRIMARY IONS FROM SIMPLE CYCLIC STRUCTURES

ALLAN B. DELFINO and ARMAND BUCHS

Department of Physical Chemistry, University of Geneva, 1211 Geneva 4, Switzerland

(Received 7 February 1973; accepted 10 October 1973)

**Abstract**—The Ion Generator computer program has now the ability to fragment cyclic structures. Examples of mechanisms which are proposed by the program for the fragmentation of simple ring compounds are presented.

IN A PREVIOUS publication<sup>1</sup> we have presented a heuristic computer program, the Ion Generator (IG), which constructs fragmentation mechanisms from acyclic molecules. The program has now been expanded in order to accept ring structures as well as acyclic structures and we present here the program's newly acquired ability to propose mechanisms for the fragmentation of cyclic structures. Cyclohexanone, 2-methylcyclohexanone and 2-methylpiperidine have been chosen as representative examples of simple ring compounds.

In order to keep the amount of output to a manageable size, the generation process is restricted for the present to the production of primary ions only. This restriction will be lifted as soon as meaningful methods of mechanism selection can be devised, but the program is able to fragment the primary ions in much the same way that it does fragment molecular ions. At the present stage of development the output contains many mechanisms which, although they are strictly correct as far as the electron-bookkeeping method is concerned, should be discarded when compared to other energetically more favourable mechanisms proposed for the same ion by the IG. If, for example, among the set of mechanisms which are proposed for a given ion, one of them leads to a structure showing resonance stabilisation, the program should be able, in the near future, to present only that mechanism.

To present fragmentation mechanisms the IG program applies to a structure, in all possible combinations, motivated electron-bookkeeping operations. As soon as a particular set of these motivated operations results in a mass change in the subject structure (here the molecular ions), the collection of operations is deemed a mechanism. The masses and structures of the resultant ionic and neutral species are then displayed along with the sequence of operations that caused them to be formed. The program does assure that all possible motivated combinations of electron-bookkeeping operations are tried, and tried only once, in a given structural environment.

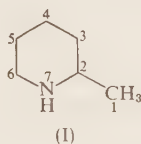
The electron-bookkeeping operations which are known to the program are indivisible units of action; they are called 'primitive operations'. The primitive operations which were implemented to give rise to the mechanisms presented here are: (1) Ionization on selected structural environments. (2) Homolytic bond cleavage

\* For Part I, see Ref. 1.

motivated by the presence of a radical site in a  $\beta$  position with respect to the fissionable bond. (3) Bond formation between neighbouring radical sites. (4) Rearrangement reactions involving the transfer of a H atom or of a group of atoms to an existing radical site through cyclic transition states of various sizes, which results in a newly located unpaired electron. (5) Intramolecular attack of a bond by a radical site followed by expulsion of a neutral species and simultaneous ring closure, which may occur in the neutral as well as in the ionic species.

### *Input to the IG program*

The data submitted to the IG program are a structure and its low resolution mass spectrum. The input of the data has been explained previously<sup>1</sup> for the mass spectra and for acyclic structures and is summarised here for ring structures only; 2-methylpiperidine is used as an example. The candidate molecule is first numbered in an arbitrary way. An ATOM CARD assigns then to each number its atom. The connections between the atoms are described through a series of BOND CARDS. To complete the input of 2-methylpiperidine numbered as shown below (I), three\* cards suffice.



| ATOM CARD  | C | C | C | C | C | C | N |   |
|------------|---|---|---|---|---|---|---|---|
| Column     | 1 | 2 | 3 | 4 | 5 | 6 | 7 |   |
| BOND CARDS |   |   |   |   |   |   |   |   |
| Column     | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| CARD No. 1 | 1 |   |   | 1 |   |   | 7 | 1 |
| CARD No. 2 | 0 |   |   | 2 |   |   | 7 |   |

The first BOND CARD transmits the following information: 'There is a group of atoms (digit '1' in column 1), from atom No. 1 to atom No. 7 inclusive, in which each atom is connected to the next in the numbering sequence 1 to 7 by a single bond (digit '1' in column 8)'. The second BOND CARD connects atom No. 2 to atom No. 7, i.e. only two atoms (digit '0' in column 1) by a single bond.

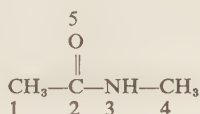
### *Output from the IG program*

As was previously mentioned<sup>1</sup> two kinds of output are presented. One is a detailed output in which each step of a mechanism is printed, along with each intermediate structure. In the other output the mechanisms are presented in an abbreviated notation. Examples of both types of output follow.

The structures of the ionic and neutral species are printed in a linear notation similar to Lederberg's DENDRAL notation,<sup>2</sup> which is used in various publications of the DENDRAL research project.<sup>3,4,5</sup> While there exists only minor differences between the notation used by the IG program and the DENDRAL notation for

\*Whatever the numbering of the structure, there is always only one ATOM CARD. The number of BOND CARDS which are needed depends on the numbering scheme, however.

representing acyclic structures, the method we use to represent rings is a departure from the DENDRAL ring notation.<sup>5</sup> Although the linear notation used by the IG program to represent acyclic structures was explained in our latest publication,<sup>1</sup> we would like to summarise again the basic rule of this notation: 'An atom is printed along with the hydrogen atoms attached to it and is followed by its set of efferent bonds printed in any order. The radicals which are connected to these bonds are then written in the same order as were the bonds'. As an example we have represented in Scheme 1 the 16 different linear notations by which *N*-methylacetamide can be represented. As can be seen from these examples, the notation can start with any one of the nonhydrogen atoms. The canonical DENDRAL notation has the advantage of resulting in but one of these 16 representations.

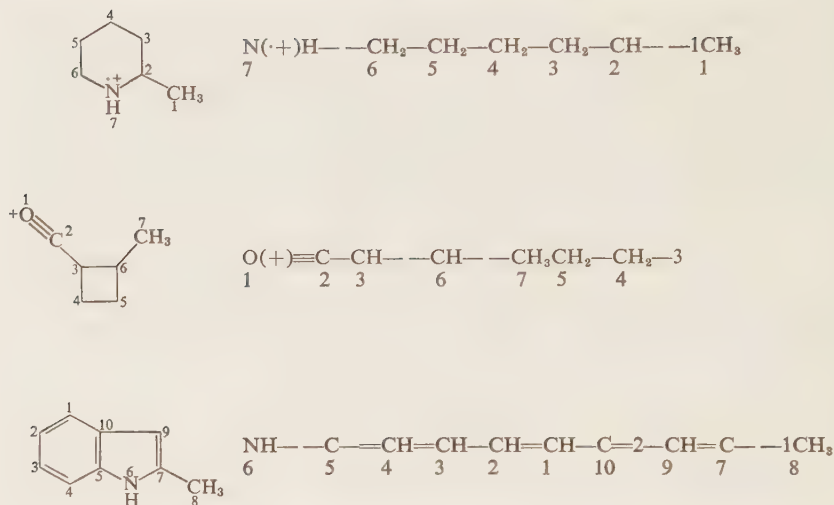


Linear notation  
starts at atom No:

|   |                                                                                                                                                                                                                                                                                                                                    |
|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | $\text{CH}_3 - \text{C} = \text{ONH} - \text{CH}_3$<br>$\text{CH}_3 - \text{C} = \text{NH} - \text{CH}_3\text{O}$                                                                                                                                                                                                                  |
| 2 | $\text{C} = \text{CH}_3\text{NH} - \text{CH}_3\text{O}$<br>$\text{C} = \text{NH} - \text{CH}_3\text{CH}_3\text{O}$<br>$\text{C} = \text{CH}_3\text{ONH} - \text{CH}_3$<br>$\text{C} = \text{NH} - \text{CH}_3\text{OCH}_3$<br>$\text{C} = \text{OCH}_3\text{NH} - \text{CH}_3$<br>$\text{C} = \text{ONH} - \text{CH}_3\text{CH}_3$ |
| 3 | $\text{NH} - \text{CH}_3\text{C} = \text{CH}_3\text{O}$<br>$\text{NH} - \text{CH}_3\text{C} = \text{OCH}_3$<br>$\text{NH} - \text{C} = \text{OCH}_3\text{CH}_3$<br>$\text{NH} - \text{C} = \text{CH}_3\text{OCH}_3$                                                                                                                |
| 4 | $\text{CH}_3 - \text{NH} - \text{C} = \text{OCH}_3$<br>$\text{CH}_3 - \text{NH} - \text{C} = \text{CH}_3\text{O}$                                                                                                                                                                                                                  |
| 5 | $\text{O} = \text{C} - \text{CH}_3\text{NH} - \text{CH}_3$<br>$\text{O} = \text{C} - \text{NH} - \text{CH}_3\text{CH}_3$                                                                                                                                                                                                           |

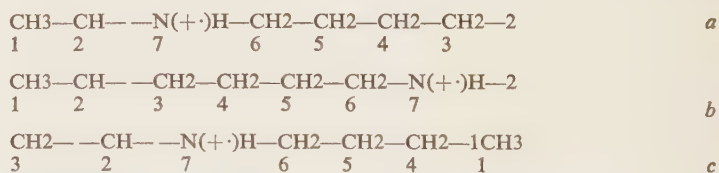
SCHEME 1. The sixteen possible linear notations for *N*-methylacetamide.

In order to represent cyclic structures, one of the bonds in the string is repeated twice. The second time the bond is printed it is followed by a number which indicates to which nonhydrogen atom in the notation string, already printed, it is connected. Scheme 2 shows examples of linear notations for three cyclic structures; the atom numbers are given as a guide to the understanding of the notation. Since a notation can start with any one of the nonhydrogen atoms and can traverse the structure in any of all the possible paths, the notations shown in Scheme 2 for each structure are not unique, but they are unambiguous. A cyclic structure, like an acyclic one, can be



SCHEME 2. Examples of structures to show linear notations of rings.

represented by many different strings of characters. For the molecular ion of 2-methylpiperidine, for example, the following are notations which are equivalent to the one already presented in Scheme 2.



In the last string (c) for example, the repeated bond, followed by the digit '1', is in this case the bond which connects atom No. 4 to the *first* atom in the string, namely atom No. 3 (and not the atom numbered one). From these rules it follows that a string has to terminate with either an acyclic radical or a number.

*Examples of mechanisms proposed for the fragmentation of cyclohexanone, 2-methyl-cyclohexanone and 2-methylpiperidine*

The IG program was run with the following choices of the available options: (1) The ionisation threshold value was set so as to ionise only the heteroatom lone-pairs (see Ref. 1). (2) In rearrangement reactions only hydrogen atoms were allowed to be transferred. (3) The sizes of the cyclic transition states allowed were restricted to 4-, 5- and 6-membered rings. (4) A maximum of two rearrangement reactions were allowed to occur in a mechanism. (5) Transfers of hydrogen atoms which are internal to the end ionic or neutral species and thus could not be substantiated by D-labelling experiments were not allowed to occur. (6) The sizes allowed for ring-closing reactions ranged from 3- to 6-membered rings.

The generation process was allowed to proceed without reference to the actual mass spectra and thus to eventually generate ions which are not in the spectra. The user has the option to restrict the ion generation process to the formation of only those ions which are supported by the mass spectrum.

The mass spectra of cyclohexanone, 2-methylcyclohexanone and 2-methylpiperidine are reported in Table 1. The ions which have been generated by the IG program are labelled with a '+' sign. With cyclohexanone, mechanisms were proposed for all the important ions (Rel. int. > 5%) found in the spectrum with the exception of  $m/e$  39, 40, 41, 54 and 80. It is known from high resolution spectra<sup>6</sup> that all these are

TABLE 1

| $m/e$ | Relative intensities (% of Base peak) |                                    |                                 |
|-------|---------------------------------------|------------------------------------|---------------------------------|
|       | Cyclohexanone <sup>a</sup>            | 2-Methylcyclohexanone <sup>b</sup> | 2-Methylpiperidine <sup>c</sup> |
| 30    |                                       | 1                                  | 10 +                            |
| 39    | 27                                    | 58                                 | 7                               |
| 40    |                                       | 13                                 | 1                               |
| 41    | 37                                    | 98                                 | 11                              |
| 42    | 90 +                                  | 68 +                               | 21                              |
| 43    | 12 +                                  | 24 +                               | 14 +                            |
| 44    | 1                                     | 1                                  | 6 +                             |
| 50    | 2                                     | 4                                  | 1                               |
| 51    | 3                                     | 6                                  | 2                               |
| 52    |                                       | 2                                  | 1                               |
| 53    | 3                                     | 10                                 | 2                               |
| 54    | 7                                     | 59                                 | 3                               |
| 55    | 100 +                                 | 83 +                               | 10                              |
| 56    | 11 +                                  | 74 +                               | 38 +                            |
| 57    | 2 +                                   | 7 +                                | 14 +                            |
| 58    |                                       | 4                                  | 3 +                             |
| 66    |                                       | 3                                  | 1                               |
| 67    | 1                                     | 8                                  | 4                               |
| 68    | 1                                     | 100                                | 3                               |
| 69    | 25 +                                  | 57 +                               | 6                               |
| 70    | 20 +                                  | 12 +                               | 6 +                             |
| 71    | 2                                     | 6 +                                | 9 +                             |
| 79    |                                       | 3                                  |                                 |
| 80    | 4                                     |                                    |                                 |
| 81    |                                       | 2                                  | 3                               |
| 82    |                                       | 1                                  | 3                               |
| 83    | 7 +                                   | 13 +                               | 1                               |
| 84    |                                       | 29 +                               | 100 +                           |
| 94    |                                       | 3                                  | 1                               |
| 97    | 2                                     | 7 +                                | 2                               |
| 98    | 31 +                                  |                                    | 10 +                            |
| 99    |                                       |                                    | 12 +                            |
| 112   |                                       | 47 +                               |                                 |

+ = Ions which were generated.

<sup>a</sup> MSDC data sheet No. 4303.

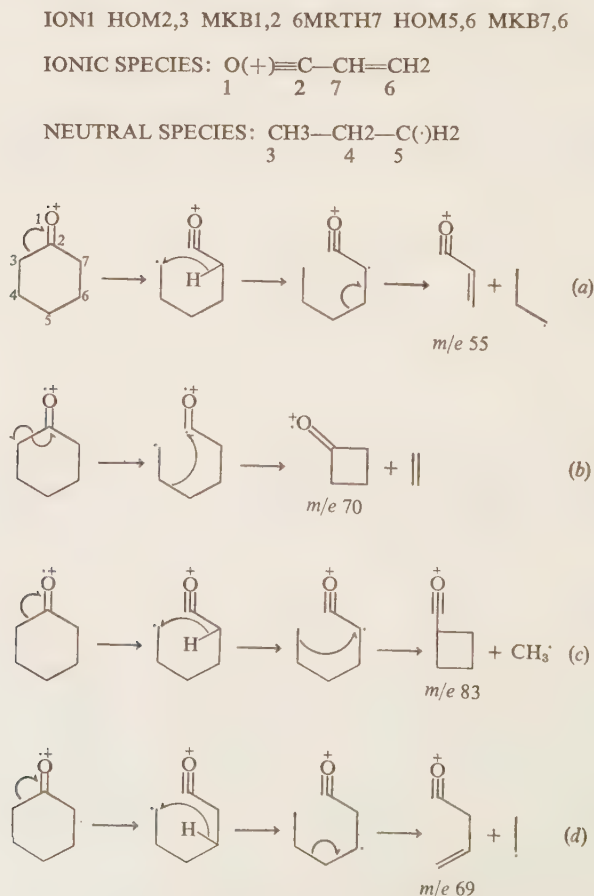
<sup>b</sup> MSDC data sheet No. 4710.

<sup>c</sup> From Ref. 6.

hydrocarbon type ions. At the present stage of development our program does not perform heterolytic cleavages, so that the charge cannot be moved away from the atom on which it was created by the action of the primitive operation 'ionisation'. The same reason accounts for the nongeneration of  $m/e$  39, 40, 41, 54 and 68 from 2-methylcyclohexanone. Again it can be seen (Table 1) that all major oxygen containing ions were generated. In the fragmentation of 2-methylpiperidine the program presented no mechanism for the formation of  $m/e$  41 and 42 which are known to contain in their structure the nitrogen atom.<sup>7</sup> The other ions which were not generated are all hydrocarbon ions ( $m/e$  55, 40 and 39). Among the numerous mechanisms which were proposed four have been chosen for each structure to illustrate the performance of the IG program.

*Cyclohexanone.* The four mechanisms are reproduced in Scheme 3.  $m/e$  70 (mechanism b).

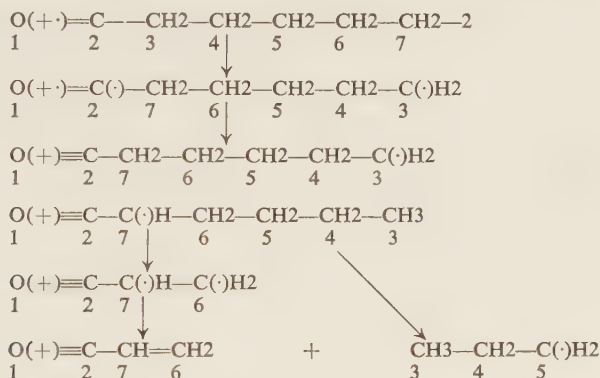
The well known mechanism which rationalises the formation of that ion is presented as follows:



SCHEME 3. Examples of mechanisms proposed for cyclohexanone.

This abbreviated mechanism notation, similar to the one printed by the program, translates as follows: Ionisation on atom No. 1, homolysis between atom No. 2 and atom No. 3, bond formation between atom No. 1 and atom No. 2, 6-membered transition state to transfer an H atom originating from atom No. 7 (transfer to the existing radical site), homolysis between atom No. 5 and atom No. 6, bond formation between atom No. 6 and atom No. 7'.

If the user requests a detailed output the following intermediate structures will be printed:



The IG program is able to recognise that the mechanism which starts with  $\alpha$ -cleavage on the other side of the molecular ion HOM2,7) is equivalent and rejects it from the output.

#### *m/e 70 (mechanism b)*

This mechanism illustrates an example of a ring-closing reaction.

ION1 HOM2,3 HOM4,5RC5,2

IONIC SPECIES:  $\text{O}(+) = \underset{1}{\text{C}} - \underset{2}{\text{CH}_2} - \underset{7}{\text{CH}_2} - \underset{6}{\text{CH}_2} - \underset{5}{\text{CH}_2} - 2$

NEUTRAL SPECIES:  $\text{C}(\cdot)\text{H}_2 - \underset{3}{\text{C}}(\cdot)\text{H}_2^*$

#### *m/e 83 (mechanism c)*

ION1 HOM2,3 MKB1,2 6MRTH7 HOM3,4RC4,7

IONIC SPECIES:  $\text{O}(+) = \underset{1}{\text{C}} - \underset{2}{\text{CH}} - \underset{7}{\text{CH}_2} - \underset{6}{\text{CH}_2} - \underset{5}{\text{CH}_2} - \underset{4}{\text{CH}_2} - 2$

NEUTRAL SPECIES:  $\text{C}(\cdot)\text{H}_3$

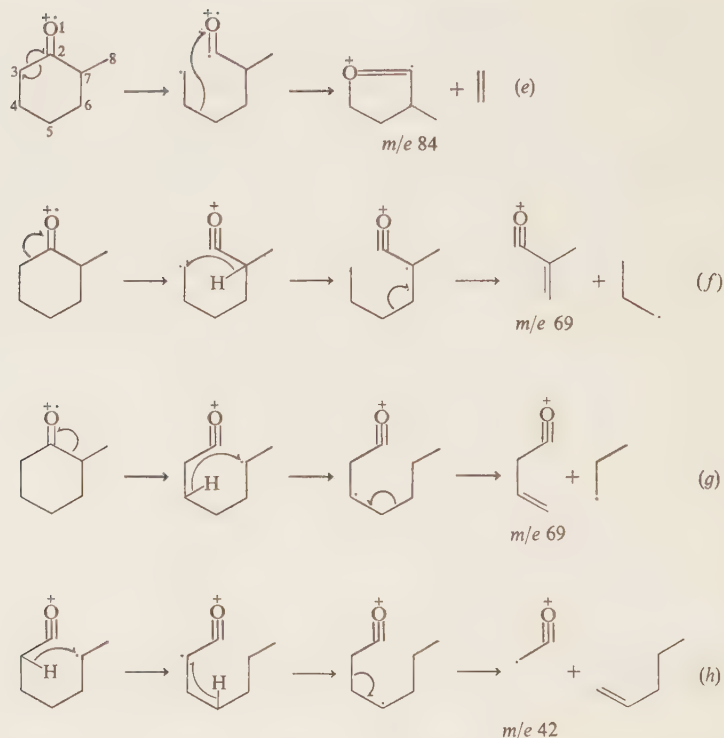
#### *m/e 69 (mechanism d)*

ION1 HOM2,3 MKB1,2 5MRTH6 HOM4,5 MKB5,6

IONIC SPECIES:  $\text{CH}_2 - \underset{7}{\text{C}} = \underset{2}{\text{O}}(+) - \underset{1}{\text{CH}} = \underset{6}{\text{CH}} - \underset{5}{\text{CH}_2}$

NEUTRAL SPECIES:  $\text{CH}_3 - \underset{3}{\text{C}}(\cdot)\text{H}_2$

\*At the present time, once the end fragmentation step has occurred, the program does not work further on the neutral species.



SCHEME 4. Examples of mechanisms proposed for 2-methylcyclohexanone.

2-Methylcyclohexanone. The four examples are given in Scheme 4.

*m/e* 84 (mechanism *e*)

An interesting example of ring closure is proposed among the many mechanisms rationalising the expulsion of ethylene.

ION1 HOM2,3 HOM4,5RC5,1

IONIC SPECIES:  $\text{O}(+) = \text{C}(\cdot) - \text{CH} - \text{CH}_2 - \text{CH}_2 - \text{CH}_3$   
 1 2 7 6 5 8

NEUTRAL SPECIES:  $\text{C}(\cdot)\text{H}_2 - \text{C}(\cdot)\text{H}_2$   
 3 4

*m/e* 69 (mechanism *f*)

This classical mechanism, which is analogous to the one yielding *m/e* 55 with cyclohexanone, is presented as follows:

ION1 HOM2,3 MKB1,2 6MRTH7 HOM5,6 MKB6,7

IONIC SPECIES:  $\text{O}(+) = \text{C} = \text{C} - \text{CH}_2\text{CH}_3$   
 1 2 7 6 8

NEUTRAL SPECIES:  $\text{CH}_3 - \text{CH}_2 - \text{C}(\cdot)\text{H}_2$   
 3 4 5

Another mechanism proposed for  $m/e$  69 is shown in Scheme 4 (mechanism g).

ION1 HOM2,7 MKB1,2 5MRTH4 HOM5,6 MKB4,5

IONIC SPECIES:  $\text{O}(+)\equiv\text{C}-\text{CH}_2-\text{CH}=\text{CH}_2$   
 $\begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix}$

NEUTRAL SPECIES:  $\text{C}(\cdot)\text{H}_2-\text{CH}_2-\text{CH}_3$   
 $\begin{matrix} 6 & 7 & 8 \end{matrix}$

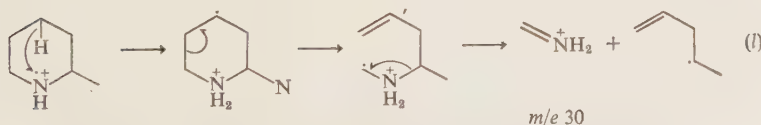
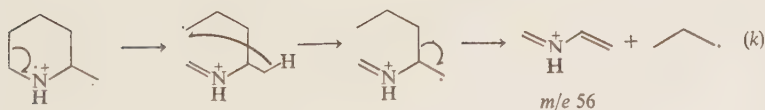
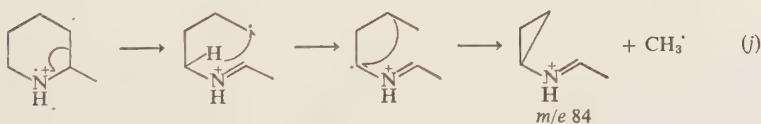
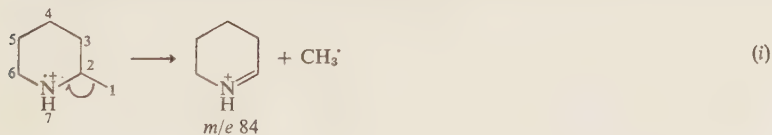
$m/e$  42 (mechanism h)

ION1 HOM2,7 MKB1,2 6MRTH3 4MRTH5 HOM3,4

IONIC SPECIES:  $\text{O}(+)\equiv\text{C}-\text{C}(\cdot)\text{H}_2$   
 $\begin{matrix} 1 & 2 & 3 \end{matrix}$

NEUTRAL SPECIES:  $\text{C}(\cdot)\text{H}_2-\text{C}(\cdot)\text{H}-\text{CH}_2-\text{CH}_2-\text{CH}_3$   
 $\begin{matrix} 4 & 5 & 6 & 7 & 8 \end{matrix}$

2-Methylpiperidine. The four mechanisms presented below are given in Scheme 5.



SCHEME 5. Examples of mechanisms proposed for 2-methylpiperidine.

$m/e$  84 (mechanism i)

ION7 HOM1,2 MKB2,7

IONIC SPECIES:  $\text{N}(+)\text{H}-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}=\text{CH}_2$   
 $\begin{matrix} 7 & 6 & 5 & 4 & 3 & 2 \end{matrix}$

NEUTRAL SPECIES:  $\text{C}(\cdot)\text{H}_3$   
 $\begin{matrix} 1 \end{matrix}$

A mechanism analogous to the one suggested by Djerassi *et al.*<sup>8</sup> for the loss of  $\text{CH}_3$  from piperidine is also proposed by the IG program for 2-methylpiperidine (mechanism *j*).

ION7 HOM2,3 MKB2,7 5MRTH6 HOM3,4RC4,6

IONIC SPECIES:  $\text{CH}_2\text{---}\underset{5}{\text{---CH}_2}\text{---}\underset{4}{\text{CH}}\text{---}\underset{6}{\text{---1N(+)H}}\text{---}\underset{7}{\text{CH}}\text{---}\underset{2}{\text{CH}}\text{---}\underset{1}{\text{CH}_3}$

NEUTRAL SPECIES:  $\text{C}(\cdot)\text{H}_3$   
3

*m/e* 56 (mechanism *k*)

ION7 HOM5,6 MKB6,7 6MRTH1 HOM2,3 MKB1,2

IONIC SPECIES:  $\text{CH}_2\text{=N(+)H}\text{---}\underset{6}{\text{CH}}\text{---}\underset{7}{\text{CH}}\text{---}\underset{2}{\text{CH}}\text{---}\underset{1}{\text{CH}_3}$

NEUTRAL SPECIES:  $\text{CH}_3\text{---}\underset{5}{\text{CH}_2}\text{---}\underset{4}{\text{C}(\cdot)\text{H}_2}$   
3

*m/e* 30 (mechanism *l*)

ION7 5MRTH4 HOM5,6 MKB5,4 HOM2,7 MKB6,7

IONIC SPECIES:  $\text{CH}_2\text{=N(+)H}_2$   
6 7

NEUTRAL SPECIES:  $\text{CH}_2\text{=CH}\text{---}\underset{5}{\text{CH}_2}\text{---}\underset{4}{\text{CH}}\text{---}\underset{3}{\text{CH}_2}\text{---}\underset{2}{\text{C}(\cdot)\text{H}}\text{---}\underset{1}{\text{CH}_3}$

Bond formation between atom No. 4 and atom No. 5 is performed despite the fact that these two atoms end up in the neutral species. At the time that bond was formed the IG program did not know what the neutral species would turn out to be. These examples show that at the present stage of development the IG computer program proposes plausible mechanisms for the fragmentation of simple ring compounds. One of the main difficulties encountered lies in the amount of output which is produced, even when the program is restricted as it was for the runs presented here. Before an attempt can be made towards the generation not restricted just to primary ions or towards handling more complex molecules, heuristic rules have to be found to select mechanisms which are not only possible but favourable. Other primitive operations still remain to be added; for example, heterolytic cleavage as mentioned previously. We feel, however, that already at the present stage of developments, if given a set of labelled compounds the IG program could prove useful in selecting among the proposed mechanisms only those which are substantiated by the mass spectra of the labelled analogues. No attempt has however been made toward such an experiment. Many developments in the handling, presentation and pruning of the output still remain to be accomplished before the program can be considered to have a real value as an aid to mass spectrometrists.

#### EXPERIMENTAL

The IG computer program runs on the CDC-3800 computer at the Computer Center of the University of Geneva. It is written in the FORTRAN programming language. As an example, two minutes of computer time were needed to fragment 2-methylpiperidine; 101 mechanisms were proposed, including those involving undetectable transfers of hydrogen atoms.

*Acknowledgements*—This work was partially supported by a grant (No. 2.648.72) received from the 'Fonds National Suisse de la Recherche Scientifique'. We also thank Mrs F. Kloeti and Miss B. May for technical assistance.

#### REFERENCES

1. Part 1. See A. B. Delfino and A. Buchs, *Helv. Chim. Acta* **55**, 2017 (1972).
2. J. Lederberg, in *Topology of Molecules*, The Mathematical Sciences, MIT Press, Cambridge, Mass., 1969, p. 37.
3. A. M. Duffield, A. V. Robertson, C. Djerassi, B. G. Buchanan, G. L. Sutherland, E. A. Feigenbaum and J. Lederberg, *J. Amer. Chem. Soc.* **91**, 2977 (1969).
4. G. Schroll, A. M. Duffield, C. Djerassi, B. G. Buchanan, G. L. Sutherland, E. A. Feigenbaum and J. Lederberg, *J. Amer. Chem. Soc.* **91**, 7440 (1969).
5. Y. M. Sheikh, A. Buchs, A. B. Delfino, G. Schroll, A. M. Duffield, C. Djerassi, B. G. Buchanan, G. L. Sutherland, E. A. Feigenbaum and J. Lederberg, *Org. Mass Spectrom.* **4**, 493 (1970).
6. J. Seibl and T. Gäumann, *Helv. Chim. Acta* **46**, 2857 (1963).
7. R. A. Saunders and A. E. Williams, in W. L. Meade, (Ed), *Advances in Mass Spectrometry*, Vol. 3, The Institute of Petroleum, London, 1966, pp. 681 to 699.
8. A. M. Duffield, H. Budzikiewicz, D. H. Williams and C. Djerassi, *J. Amer. Chem. Soc.* **87**, 810 (1965).

Appendix D.

This paper was presented at the International Computing Symposium, 4-7 September 1973 at Davos, Switzerland. The paper thereafter appeared in the proceedings published by the North-Holland Publishing Company during 1974. The paper dealt with the computer science aspects of the ION GENERATOR.

## HEURISTIC PROGRAMMING TECHNIQUES APPLIED TO CHEMICAL MASS SPECTROMETRY

Allen Delfino and Bernard Levrat

Département de Chimie-Physique et Centre interfacultaire  
de calcul électronique de l'Université de Genève

A program is presented which searches out theoretical explanations for the appearance of ions in the mass spectra of organic compounds. The techniques of heuristic search and recursive programming are applied on appropriate data structures. Heuristic techniques are applied to accomplish forward pruning to stop the generation of chemically implausible ions.

### 1. INTRODUCTION

Chemical mass spectrometry is concerned with the products of a device which gives a large dose of internal energy to an unknown substance, causing it to undergo disintegration reactions, resulting in a set of reproducible ionic fragments whose mass and relative intensities can be measured with great accuracy. Chemists use this pattern to work back to the structure of the original unknown structure. But it was rapidly found that the chemical mechanisms for the occurrence of these fragments would need to be understood, to be able to solve larger structures of chemical and biological interest. Hence research was launched to elucidate these mechanisms by disintegrating known compounds and rationalizing the results into a unifying theory. Armed with this theory, various workers in the field have written programs to solve unknown substances [1-5]. This work tries to predict the resulting ions in the mass spectrum from the known structure by applying accepted mechanistic steps via a computer program. If an important ion turns out to be missing, this will induce the chemist to suggest new mechanistic steps to arrive at the given mass.

The ION GENERATOR program has been implemented in FORTRAN on the CDC 3800 of the University of Geneva to carry out rationalizations of mass spectra in a methodical, exhaustive manner. By having programmed all accepted theory up to the study being undertaken, it can be immediately determined whether or not new principles are needed to explain the appearance of the ions in the spectra of the known compounds being studied.

Advanced computing techniques were found necessary to carry out this task. An ap-

propriate data structure was applied to represent the molecular structure. On arriving at a fragmentation mechanism, one must search for the intensity recorded at the mass of the fragment ion. Therefore the search time for finding an ion in the spectrum was optimized.

The heart of the program, the generation of the mechanisms from the various motivated reaction steps, is a recursive algorithm, similar to the exhaustive move generators of game playing programs. In fact, one can define mechanism, in BNF fashion [5], as :

```
<mechanism> ::= <reaction step> |  
                <reaction step> <mechanism>  
  
<reaction step> ::= ionization |  
                    hydrogen migration | ... | homolysis
```

On any given structure, perhaps several reaction steps can be done. Therefore it is necessary to determine these and to store them away on a push-down stack for execution at a later time, after the full depth of the first of the reactions has been explored. Then each, in their turn, can be explored in full depth, as was the first.

When a reaction step results in a fragmentation of a structure, one looks in the spectrum for the intensity of the fragment, which might have been recorded in the actual mass spectrum. But at that point, the reaction steps leading to the structure need to be noted, and an estimation of the stabilities of the products needs to be made to evaluate the mechanism a posteriori. A stack of records, called the "diary", was implemented to keep track of the sequence of mechanistic steps. These records are used afterwards to display the mecha-

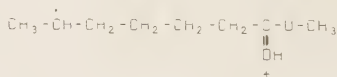
The chemist user will quickly notice that a mass from the spectrum was missed, i.e. not generated by any sequence of the reactions. It is then up to the chemist to suggest a new reaction to be included in the mechanisms proposed by the program for all future compounds.

Cyclic or ring structures are of special interest in mass spectrometry. Compounds also are thought to enter into rings during the transition from one structure to another; therefore a special data structure called the "chemical ring" memory was implemented to decrease the time needed to manipulate ring-containing compounds which also will form these transitory rings. With the implementation of this data structure, the range of compounds, on which one could feasibly run the program, was significantly increased. In fact, the system accepts compounds of any complexity containing up to fifty atoms. This does not mean that every ion will be generated that has ever appeared in a mass spectrum, but only that the potential to do so is there. Hence it should be a useful tool for chemists.

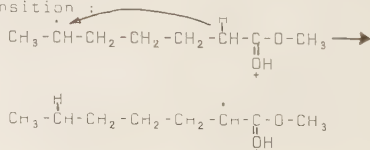
The system includes a formatted memory dump of all program data areas which has been of great value in debugging the program. The creation of diverse data structures demands that their contents be interpreted in terms of their uses when the program detects an inconsistency. An octal dump of memory is meaningless in a program which is operating on higher level data. The bits in a machine are only the medium, and not the information.

The heart of the ION GENERATOR program from the chemical point of view and from the heuristic programming point of view is the recursive mechanism generator, along with the heuristic devices used for pruning the massive number of branches of the tree of mechanisms which results from using the generator. As was pointed out in the introductory remarks, mechanisms are made up of a series of reaction steps which operate on structures one after another until a fragment is formed. This can be thought of as a tree, like the one pictured in Figure 1, where each node represents a structure and each efferent arrow, a reaction on the nodal structure. The topmost node is the neutral molecule which is to undergo two ionization reactions in this example. (Ionization reactions can only occur once on the neutral molecule, hence these reactions never occur deeper in the tree.) At the second level starts the application of reactions on the molecular ions formed by ionization. The reactions which occur at this level can occur over and over until heuristic pruning stops the process.

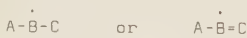
Every reaction must find the proper environment that it needs to be motivated or it will not be considered. The reaction of hydrogen migration via a cyclic transition state will be used to illustrate how a process is determined to be possible. Given the following structure:



which resulted from an 8 member transition state (first migration) from the site of the oot to the oxygen where the charge is residing (+). If another transition state is to take place now that one migration has already happened, the program insures that the resulting radical site environment must give rise to resonance stabilization. In this particular ion, this can occur only if one considers the following transition :



Chemical considerations thus limit the possibilities for a second hydrogen migration, which are numerous. Other heuristic devices work at reducing the possibilities for the first migration. For example, the following situations are not permitted to occur ever :



Also once a hydrogen has been moved, it cannot be moved again in the same mechanism.

For a molecule of 14 atoms, as in the example given for the final results, one can consider that the maximum depth of levels which can occur is 6; since there are 40 different bonds that can break, one can use the combinations formula to calculate how many ions are possible. The value is approximately  $10^6$ . And the number of mechanisms to get them is perhaps as many as ten times that because often vastly differing mechanisms can result in the same ion. So the heuristic constraints are the crux of this problem, like so many of the other problems programmed with recursive generators.

If a reaction is found to be unfavorable for one reason or another, the effect is that a complete branch of the mechanism tree will be avoided or pruned. It would be inappropriate in an exposition such as this to delve into all of the chemical considerations which are used to avoid unwanted mechanisms. But we would like to point out that all sequences of reactions are possible in theory, but when a particular structure is presented certain reactions can be ruled out.

Certain pruning criteria are allowed in consideration of the fashion in which mass spectrometry is used. If mass spectrometry is to be efficacious as a structure elucidation method, reactions which rearrange the skeletal structure severely should not be considered. In other words, the least rearranging mechanism will be preferred. Also, the smaller the ions become, the more ways they can usually be formed. Hence, the size of an ion with respect to the original molecule could trigger the decision to stop the search for ion producing mechanisms.

### 3. PUSH-DOWN STACKS FOR NODES

All the allowed reactions, which can happen to a structure represented by a node in the mechanism generation tree, are determined at one time, but can not take place all at once. So, the information determined for a given node is saved for later use, in turn, on a push-down stack allocated especially for that node. Since the number and size of each stack is unknown before-hand, a system of stack allocation, access and storage was implemented.

First of all, an available space for stacks of 2000 words was allocated in a COMMON block. Then 3 subroutines were implemented for the allocation of a stack for a node, the storage (push-down) of an element, and the retrieval of a node (pop-up). So that, in terms of the tree, each set of efferent arrows from a node is equivalent to a stack, and when the end-of-stack occurs for that node, all the reactions will have been tried, and return will then be made to the prior node, to which the now finished node is attached (see Figure 2).

The push-down stacks for the nodes are not the only stacks required by the system. The subroutine which creates the linear notation described in Delfino and Buchs [6], needs a stack, as does the structure traversal routine. These routines, in essence, borrow from the unused portion of the space left available by the stack system on a strictly temporary basis. But they do this outside of the stack allocation sub-system.

### 4. DIARY STACK

The diary stack is a large stack of 30 elements per entry, in which every item of importance for an executed reaction step is stored, along with other useful information. At any time, the stack represents the mechanism which transformed the given neutral molecule to the current ionic species, which is about to undergo further reactions and possibly further fragmentations. Two significant uses are made of this stack.

The first use is when a structure fragments, a recording is made on an output tape for the Phase II display. At this time, an evaluation of stability is made. This estimate is made in an absolute way, without regard for other mechanisms, and

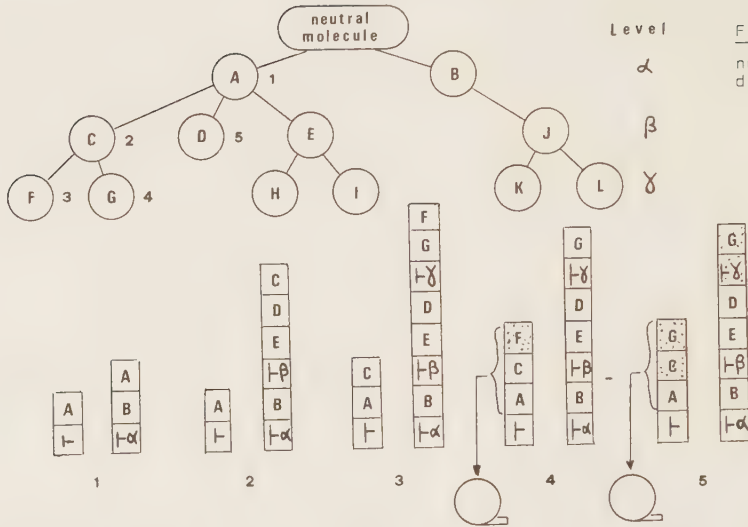


Figure 1 :  
node and  
diary stacks

| I  | II | III | I  | II  | III | I   | II | III | I   | II | III |
|----|----|-----|----|-----|-----|-----|----|-----|-----|----|-----|
| 14 | 3  | 15  | 60 | 0   | 5   | 95  | 1  | 0   | 135 | 1  | 0   |
| 15 | 2  | 154 | 61 | 0   | 3   | 96  | 2  | 0   | 137 | 2  | 0   |
| 16 | 6  | 0   | 65 | 1   | 1   | 97  | 7  | 2   | 139 | 5  | 0   |
| 18 | 1  | 0   | 67 | 2   | 0   | 98  | 1  | 4   | 140 | 1  | 0   |
| 26 | 1  | 0   | 68 | 1   | 2   | 99  | 2  | 14  | 141 | 1  | 0   |
| 27 | 5  | 0   | 69 | 28  | 0   | 100 | 0  | 3   | 142 | 0  | 6   |
| 28 | 3  | 4   | 70 | 5   | 4   | 101 | 73 | 8   | 143 | 2  | 1   |
| 29 | 12 | 31  | 71 | 13  | 25  | 102 | 6  | 6   | 144 | 0  | 2   |
| 31 | 1  | 4   | 72 | 1   | 0   | 103 | 1  | 3   | 145 | 1  | 0   |
| 39 | 6  | 0   | 73 | 8   | 0   | 109 | 1  | 0   | 151 | 3  | 0   |
| 40 | 1  | 2   | 74 | 8   | 0   | 110 | 2  | 0   | 153 | 3  | 4   |
| 41 | 22 | 0   | 75 | 1   | 0   | 111 | 10 | 0   | 155 | 1  | 0   |
| 42 | 6  | 0   | 77 | 1   | 0   | 112 | 2  | 9   | 157 | 7  | 0   |
| 43 | 14 | 2   | 79 | 1   | 0   | 113 | 9  | 8   | 158 | 1  | 0   |
| 44 | 1  | 0   | 81 | 2   | 0   | 114 | 1  | 0   | 167 | 1  | 7   |
| 45 | 1  | 0   | 82 | 1   | 2   | 115 | 3  | 0   | 168 | 0  | 6   |
| 51 | 1  | 0   | 83 | 13  | 0   | 121 | 1  | 0   | 169 | 8  | 9   |
| 53 | 3  | 0   | 84 | 1   | 0   | 125 | 3  | 0   | 170 | 1  | 6   |
| 54 | 1  | 4   | 85 | 4   | 0   | 126 | 2  | 0   | 171 | 3  | 1   |
| 55 | 22 | 2   | 86 | 0   | 1   | 127 | 0  | 2   | 172 | 0  | 1   |
| 56 | 6  | 2   | 87 | 21  | 5   | 128 | 0  | 3   | 182 | 0  | 5   |
| 57 | 31 | 27  | 88 | 100 | 8   | 129 | 15 | 6   | 185 | 1  | 14  |
| 58 | 2  | 1   | 89 | 5   | 6   | 130 | 1  | 3   | 200 | 5  | 1   |
| 59 | 12 | 1   |    |     |     | 131 | 0  | 4   |     |    |     |

Table I : Results of the ION GENERATOR program for  $\text{CH}_3\text{-CH}_2\text{-CH-CH}_2\text{-CH-CH}_2\text{-CH-C-O-CH}_3$   
 $\text{CH}_3 \quad \text{CH}_3 \quad \text{CH}_3\text{O}$   
 Column I contains the mass, II, the intensities and III, the number of mechanisms.

is based on the complex problems that they normally face. They are : degree of resonance stabilization, double bond conjugation, carbonium ion stability, etc.

The second major use of the diary entries is to "unreact" the current nodal structure, back to the preceding nodal structure. Hence, for each program which causes a certain reaction, there is another program which does the exact opposite with the same information which was taken from the node stack but is taken now from the last diary entry; i.e., each reaction has an inverse reaction.

For example, in figure 1, node F had to be "unreacted" in order to arrive at node G. The reaction information to get from the nodal structure C to structure G was stored for future use on the level ii push-down stack. After the "unreaction" of node F to node C, node C can undergo the pending reaction to node G.

Besides the information concerning the reaction steps, the program also stores the following things in the diary stack : ionic species mass, neutral species mass, the atoms which start the displays of the ionic species and the neutral species, the length of the transition state, the intensity of the ion in the actual mass spectrum (if the step resulted in fragmentation), the number of steps in the mechanism, the chronological sequence number of the mechanism in the first phase, the mass defect of the ion for high resolution studies, the seven measures of stability which are made on ions and the corresponding neutral species, the apparent mass of the metastable transition that would be associated, indicators of permissible syntactic order of mechanistic steps since fragmentation occurred last : thus the 30 words needed for every reaction.

## 5. STORAGE OF THE MASS SPECTRUM

Two methods of storage are used for the recorded mass spectrum, which is a table of mass-intensity pairs. The methods are either sequential with a binary search algorithm, or random with a direct search. One of the two methods is chosen by the program when the spectrum appears in the data stream.

The hashing algorithm produces an address M which is simply the modulo of the mass with respect to the size of the table (397), necessarily a prime number. Since the program has a capacity for compounds

of 50 non-hydrogen atoms, the mass spectrum can go up to approximately mass 800. Hence a collision accommodation, or rehashing algorithm had to be incorporated. A linear method, using the prime number 23, was chosen since 23 rarely occurs as an interval between mass spectral peaks. The new address, M', is calculated as follows :

$$M' = \text{MOD}(M+23, 397).$$

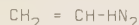
While filling the table, the program keeps track of the maximum number of collisions,  $W_R$ , which is compared to the worst possible case for the binary search

$$W_B = \text{ceiling}(\log_2 L)$$

where L is the number of peaks of the given spectrum. If  $W_R$  surpasses  $W_B$ , the hash table will be sorted and will become the binary search table. The spectrum search subroutine will then use the appropriate search algorithm.

## 6. REPRESENTATION OF MOLECULAR STRUCTURE

The representation of a molecular structure in the computer memory is accomplished via an interconnection matrix. An interconnection matrix has a row and a column for each atom in the structure. The off-diagonal elements,  $a_{ij}$ , are used to indicate bonds between the atom i and the atom j. If the number stored there is 1, a single bond is indicated; 2, a double bond, etc. The diagonal elements,  $a_{ii}$ , which would normally have no meaning in such matrices (the i<sup>th</sup> atom bonded to itself), are used to store the number of hydrogens required by the valency of the atom. These hydrogens are termed the "implicit" hydrogens. The matrix correspondent to the ethylene amine



will be

|   |   |   |
|---|---|---|
| 2 | 2 | 0 |
| 2 | 1 | 1 |
| 0 | 1 | 2 |

if atom 1 is taken to be the carbon atom with two hydrogens, atom 2 is the other carbon atom, and atom 3, the nitrogen atom. Due to the arbitrary numbering scheme, there are  $n!$  possible matrices for an n-atom structure. The program allows the user to specify the matrix at data input time and it makes no difference in the execution time which one of the  $n!$  representations is chosen; also

the resulting mechanisms are in every way equivalent.

Along with the structure go certain indicators which are also stored in matrices; these are indicators concerning the bonds of the structure. During the course of a mechanism a bond will be formed by the program, and henceforth in that sequence, that bond cannot become broken by the program. If that were allowed, it is evident that an infinite loop would result. The same is true for bonds broken by the program; they may not be reformed by a subsequent step in the same mechanism. If only on/off logical indications are needed, matrices of logical variables are used to save space.

Other properties are required by the program with regard to the structure, but they are concerned with the atoms themselves, not the bonds. These properties can be represented as vectors. For example, radical sites (atoms on which unpaired electrons are found) and electron deficient centers are properties of atoms which motivate reactions. Also an atom can enter into the same sort of reactions as do the implicit hydrogens, and hence will be given the property of "hydrogen equivalent". The atom's identity in terms of the program's periodic table is a key to all the properties which are stored there, for the atoms of the given structure. Hence, if nitrogen is one of the atoms in the structure, the program determines that it is an electron pair donor by searching the periodic table via the index to this table stored in the atomic property vector associated with the structure.

## 7. CHEMICAL RING MEMORY

Since some reactions occur, in the mass spectrometer, only in the presence of cyclic molecular (or ionic) structures, a special data structure was devised which permits the programs to ascertain, via subroutine calls, the existence of cyclic sub-structures. This data structure has been termed the "chemical ring memory". This memory has enabled the program to determine if a given transition state is flexible enough to be formed. The flexibility reduction of structures, in which rings have formed, can also be accommodated by the program.

The chemical ring memory system records all of the circuits in the neutral mole-

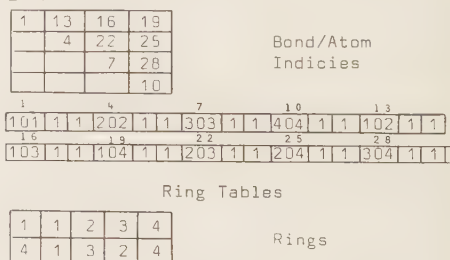
cule in a complete manner, and from then on, the reaction and unreaction subroutines inform the system of the bonds formed or broken in the course of its operation. This then, enables the system to determine quickly, if a transformation caused a change with respect to the rings in the compound. If a change did occur, the necessary updates are made at that time. This passive system gives the program the option to be efficient if the bond broken is known not to involve rings.

The chemical ring memory consists of three levels: the bond/atom indicies which point to the ring tables, which, in turn, point to the rings themselves. The bond/atom indicies are stored in a fixed-length logical matrix, in a packed mode. The ring tables are variable in number and in length, so they are allocated on demand by a sub-system, and the available space for them is garbage collected when unused space runs out. The atom number sets, which are the rings, are also variable in number and in length, hence are also garbage collected. This memory is similar to that used by Corey and Petersson [7].

The structure of a portion of the chemical ring memory will be given with cyclobutane,



stored therein. The diagram given below will be explained in the following paragraphs.



Starting at the lowest level, Rings, one finds, for the four-membered ring, five spaces used on the two-row matrix. The first position of the first or top row is the ring reference indicator. If it is non-zero, the ring is in use. If it is zero, the ring can be garbage-collected if need be. The first position of the bottom row is reserved for the length of the ring. The remaining positions of the top row are for the canonical sequence of the ring, which is a list of the atom numbers in ascending numerical order. The remaining positions of the bottom row are for the atom numbers listed in the sequence in which they are found in the circuit.

In order that the ring will be stored only one for each atom, a ring table is allocated for each bond that makes up part of the ring. If a bond is part of several rings, which is quite common, it will still have only one table; the table will just be longer. In position one is stored the bond that owns the table. If this number is zero, the table can be garbage collected. The second position of each one, is the number of rings that are stored therein. Following that is (are) stored the position(s) in the ring-space of the ring(s). In the example, all atoms and bonds have but one ring, that ring stored at position number one of the ring space.

The bond/atom index matrix has a non-zero value if the bond or atom is part of a ring. The value will be the position in the ring-table space where the ring-table will be found.

#### 8. PHASE II DISPLAY

As was mentioned in the Diary Stack section, a series of tape records, one for every fragmentation, is recorded in the Phase I of the program. At the end of the generation of all the plausible mechanisms, control is given over to the CDC SORT subsystem, on order to sort the diary records into order on the mass, plausibility, ionic formula, and type of mechanism. With these data, the Phase II program prints out each mechanism in full, those of highest mass coming out last.

The structure of methyl 2, 4, 6 trimethyl-octanoate (an ester) was given to the ION GENERATOR program. The program found 449 fragmentations. Many of the mechanisms arrive not only at the same masses as other mechanisms, but even to the same

structures. All of the mechanisms, which mass spectrometrists agree are diagnostic for esters, are found among the 449 mechanisms. If a complete set of deuterium-labelled analogs were available for this compound, choices could indeed be made among the results for a given mass.

Table 1 summarizes the results for the run, which used approximately 5 minutes of computer time. A zero in the "spectral intensity" column indicates that the program predicts a peak which was not produced in the actual scientific experiment resulted in an ion which the program with its current theory base cannot generate. The spectrum was taken from a catalog of mass spectra [8]. The interesting feature of the table is the pseudo-mass spectrum of the number of times a particular mass occurred. The pattern shows a large number of ways to arrive at small mass ions, and few ways to get the large mass ions. This shows dramatically the unambiguity of the high mass scale of a mass spectrum. The power of mass spectrometry as an analytical tool lies therein.

#### 9. BIBLIOGRAPHY

1. Crawford, L.R.; Morrison, J.D. : *Analytical Chemistry* **43**, 1790 (1971).
2. Venkataraghavan, R.; McLafferty, F.W.; Van Lear, G.E. : *Organic Mass Spectrometry* **2**, 1 (1969).
3. Buchs, A.; Delfino, A.B.; Duffield, A.M.; Djerassi, C.; Buchanan, B.G.; Feigenbaum, E.A.; Lederberg, J. : *Helvetica Chimica Acta* **53**, 1394 (1970).
4. Jurs, P.C. : *Analytical Chemistry* **43**, 1812 (1971).
5. Delfino, A.B.; Buchs, A. : *Topics in current Chemistry - Mass Spectra and Computer*, pp. 109-137 (1973), Springer-Verlag, Berlin.
6. Delfino, A.B.; Buchs, A. : *Helvetica Chimica Acta* **55**, 2017 (1972).
7. Corey, E.J.; Petersson, G.A. : *Journal of the American Chemical Society* **94**, 460 (1972).
8. Mass Spectrum of methyl 2,4,6 - trimethyl-octanoate, p. 1417, vol. II, Atlas of Mass Spectral Data, Stenham, E.; Abrahamsson, S.; McLafferty, F.W. Wiley Interscience (1969), New York.

Appendix E.

This manuscript has been submitted for publication to the *Helvetica Chimica Acta*. This article discusses the ION GENERATOR's results in the exploration of acyclic structures for ions of any generation, primary, secondary, etc.

Heuristic Programming as an Ion Generator in Mass Spectrometry. Part III.  
Fragmentation of Aliphatic Compounds

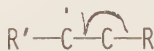
by Allan B. Delfino and Armand Buchs

Department of Physical Chemistry  
University of Geneva, 1211 Geneva 4,  
Switzerland

Summary. The computer program, known as the ION GENERATOR, is presented as an approach toward the development of a program to predict fragmentation schemes. Results are presented for various acyclic compounds.

In previous papers /1,2,3/, we have described a computer program, the ION GENERATOR (IG), which was developed to rationalize ion formations on the basis of the electron-bookkeeping theory. Up to now, our strategy was to restrict the IG program to the generation of primary ions and the aim was to arrive at a complete list of plausible fragmentation mechanisms. In this paper, we present a new strategy which allows the program to fragment not only molecular ions, as was done before, but also any daughter ion, while not generating thousands of mechanisms in the process. Basically, the program functions as it did before, but now a much stricter look is taken at the motivating structural environments before a primitive reaction step is done. Also, the end products of the reaction steps are controlled to limit the motivation of succeeding reactions. In order to achieve this degree of control, and to still generate the required ions, new reactions were defined for inclusion in the set of primitive operations. These changes result in the generation of only the chemically most favorable mechanisms; a high degree of agreement between the generated mass-only mass spectrum and the actual mass spectrum is nevertheless retained. The IG program *does not* act in a specific mode for a given structure; it does not attempt to recognize functional groups in order to apply special case reaction steps. The IG operates *in exactly the same way* whatever happens to be the candidate structure. Each mechanism is constructed *de novo*, by applying, at each node of the mechanism generation tree, all of the structurally motivated, primitive reaction steps.

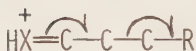
The primitive reaction steps, used by the IG program, can be classified broadly into three categories<sup>1</sup>: homolytic cleavages, heterolytic cleavages and rearrangement reactions. The formation of a new bond in the ionic or neutral species is the motivation for homolytic cleavages to be effected (not involving rearrangement). Some types of environments which trigger simple or concerted homolytic cleavages, are shown in Scheme 1.



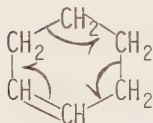
$\alpha$ -CLEAVAGE



RING CLOSURE



OVERBOND RELIEF



RETRO - DIELS - ALDER

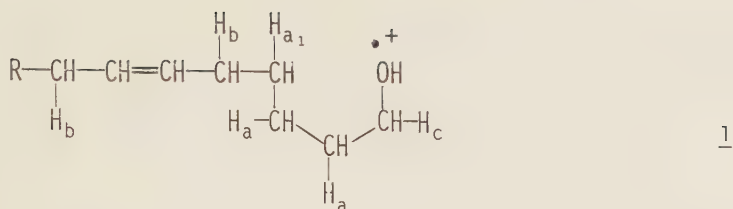
SCHEME 1.

Heterolytic cleavages are performed whenever a heteroatom bears the positive charge. Moreover, when the positive charge is located on a carbon atom with 6 electrons in its outer shell, heterolytic  $\alpha$ -cleavage and ring formations in the neutral species occur.

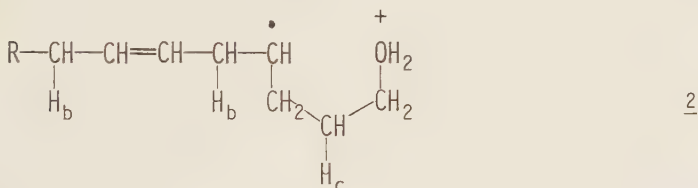
Three types of rearrangement reactions are considered.

- a) *Transfer reactions of H atoms which are caused only by the radical site, resulting in a shift in position of this site.* The choice of the H atom to be moved depends on various conditions. The ring sizes of the involved transition states are controlled. Allylic resonance is an example of the various stabilizing influences which play a role in these reactions. This is illustrated using structure 1 shown below.

<sup>1</sup> The primitive reaction step involving the ionization of the molecular structure has been explained previously /1/ and remains unchanged.



Those H atoms marked  $\text{H}_a$  and  $\text{H}_{a_1}$  are candidates for a transfer to the oxygen atom because the ring sizes that are involved are allowed (4-, 5- and 6-membered transition state rings). The  $\text{H}_b$  atoms are candidates because their transfer creates a radical site on an allylic position. When two H transfer reactions take place in the same mechanism, the second one can only affect an atom whose shift favors stabilization. If, for example, in structure 1, the transferred H atom is  $\text{H}_{a_1}$ , only the hydrogen atoms marked  $\text{H}_b$  can be shifted in the resulting structure 2.

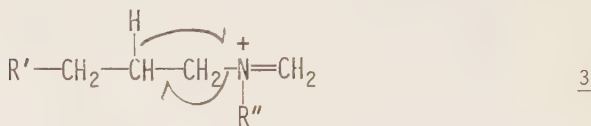


$\text{H}_c$  will not be transferred, despite the fact that the transition state is of allowed state. If during the first step (in structure 1), one of the two atoms marked  $\text{H}_b$  had migrated instead of one of the  $\text{H}_a$ 's, no further transfer would have occurred. When no particularly favorable environment triggers the transfer of an H atom, the transition state ring size cannot exceed six. This restriction is canceled however, when an H atom is favorably located, like the ones marked  $\text{H}_b$  in structure 1. Strained transition states are of course disallowed.

- b) *Transfer reactions of H atoms which are triggered by a double bond.* The classical McLafferty mechanism is included in this type of rearrangement, despite the fact that it could be classified among the radical site triggered H transfer reactions. This particular rearrangement always involves a six-membered transition state and two additional homolytic scissions.

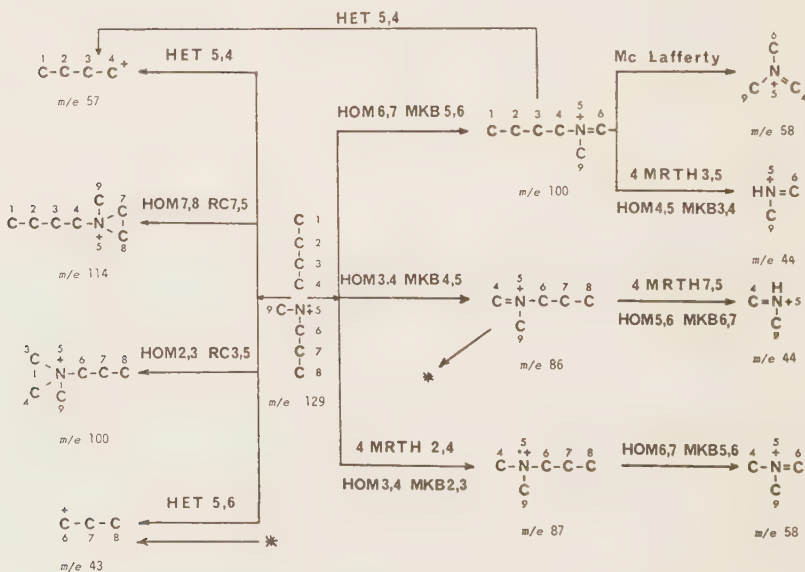
c) Transfer reactions of H atoms to overbonded heteroatoms bearing the charge.

In this case, a concerted reaction occurs, as is shown below in structure 3, depicting a classical example of this rearrangement.



The H atoms are rearranged only via four-membered transition states in this type of rearrangement, although it is known that this kind of H transfer is not site specific.

To illustrate the performance of the IG, using this new strategy, the results, obtained with four aliphatic structures, will be discussed: methyl-n-propyl-n-butylamine, ethyl sec-butyl ether, octane-3-ol, and dodecane-6-one.



SCHEME 2.

Scheme 2 shows the complete breakdown pattern which is proposed for

methyl-n-propyl-n-butylamine<sup>2</sup>. All of the classical mechanisms are among the 14 which are produced by the IG program. An examination of the 70 eV low-resolution mass spectrum of this amine (Figure 1) shows that nearly all of the diagnostically important ions are generated.

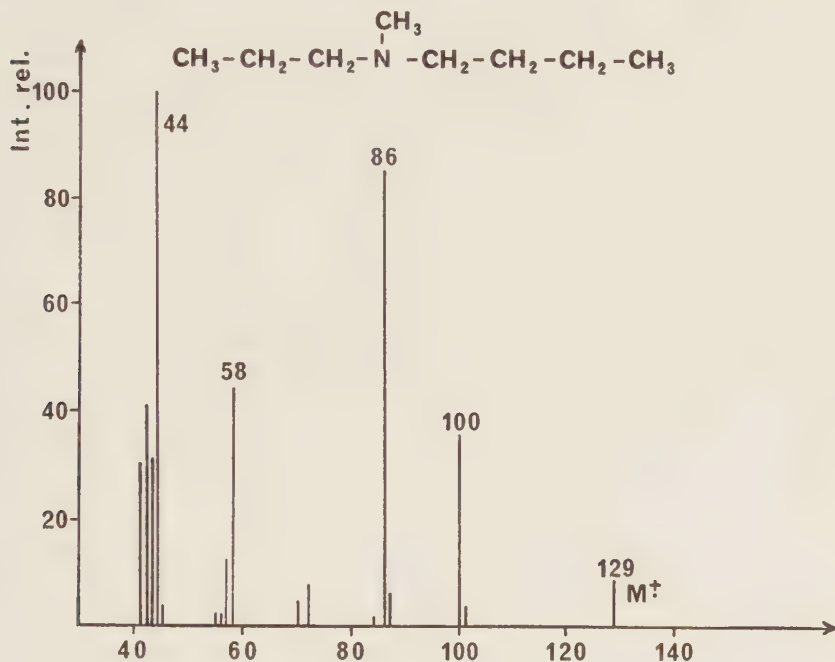


FIGURE 1.

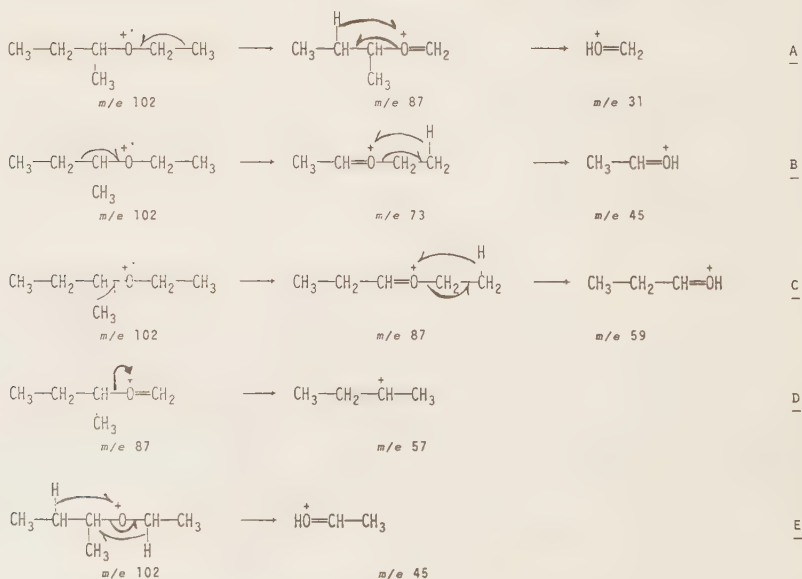
In the case of ethyl sec-butyl ether, fifteen mechanisms, including the molecular ion, are generated, leading to nine different masses in the mass-only mass spectrum. Again this strategy permits the key ions to be proposed. The ions  $m/e$  72, 55, 43, 42 and 41 are missing from the output. Table 1 shows the actual mass spectrum of ethyl sec-butyl ether and the number of times that each ion was generated. A few examples of the mechanisms, which are proposed, are illustrated in Scheme 3. Only mechanism e of scheme 3 can be considered out

<sup>2</sup> The abbreviations used in scheme 2 have been explained previously /1,2,3/.

of the ordinary because of its two consecutive hydrogen rearrangements. Otherwise, the mechanisms are highly classical in nature.

TABLE 1.  
Mass Spectrum (70 eV) of Ethyl sec-Butyl Ether.

| <i>m/e</i> | Rel. Int. (%) | Frequency | <i>m/e</i> | Rel. Int. (%) | Frequency |
|------------|---------------|-----------|------------|---------------|-----------|
| 41         | 12            | 0         | 58         | 1             | 0         |
| 42         | 2             | 0         | 59         | 19            | 1         |
| 43         | 10            | 0         | 71         | 1             | 0         |
| 45         | 100           | 3         | 72         | 2             | 0         |
| 46         | 2             | 1         | 73         | 51            | 2         |
| 55         | 3             | 0         | 87         | 4             | 4         |
| 56         | 3             | 1         | 101        | 1             | 0         |
| 57         | 10            | 2         | 102        | 1             | 1         |



SCHEME 3.

The results, obtained with octane-3-ol, are shown in Table 2 and those those for dodecane-6-one are shown in Table 3; both table indicate the actual 70 eV mass spectrum and, for each  $m/e$  value, the number of mechanisms and the number of resulting ionic structures which differ by their non-hydrogen atoms. Eighty-three mechanisms are proposed for octane-3-ol, whereas one hundred thirty-nine result with dodecane-6-one. For both compounds, the ions generated account for a large portion of the ion current. Again, all of the classical fragmentation mechanisms can be found among the results, such as  $\alpha$ -cleavages, McLafferty rearrangements, the losses of the elements of water, to name just a few.

TABLE 2.

Mass Spectrum of Octane-3-ol /4/ (70 eV)

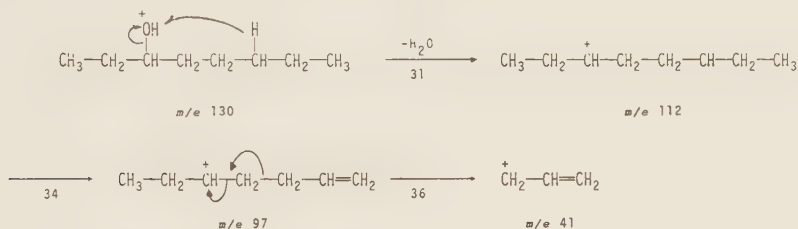
| $m/e$                    | %     | <i>M</i> | <i>F</i> | Formula                         | $m/e$                      | %    | <i>M</i> | <i>F</i> | Formula                          |
|--------------------------|-------|----------|----------|---------------------------------|----------------------------|------|----------|----------|----------------------------------|
| 41                       | 36.5  | 7        | 3        | C <sub>3</sub> H <sub>5</sub>   | 73                         | 0.5  | 4        | 1        | C <sub>4</sub> H <sub>9</sub> O  |
| 42                       | 6.8   | 1        | 1        | C <sub>3</sub> H <sub>6</sub>   | 74                         | 0.0  | 1        | 1        | C <sub>4</sub> H <sub>10</sub> O |
| 43                       | 18.2  | 2        | 1        | C <sub>3</sub> H <sub>7</sub>   | 83                         | 41.4 | 5        | 3        | C <sub>6</sub> H <sub>11</sub>   |
| 44                       | 9.3   | 0        | 0        |                                 | 84                         | 3.5  | 0        | 0        |                                  |
| 45                       | 5.4   | 2        | 1        | C <sub>2</sub> H <sub>5</sub> O | 85                         | 0.4  | 2        | 2        | C <sub>6</sub> H <sub>13</sub>   |
| 53                       | 2.2   | 0        | 0        |                                 | 87                         | 0.0  | 3        | 1        | C <sub>5</sub> H <sub>11</sub> O |
| 54                       | 1.0   | 0        | 0        |                                 | 88                         | 0.0  | 1        | 1        | C <sub>5</sub> H <sub>12</sub> O |
| 55                       | 42.3  | 3        | 2        | C <sub>4</sub> H <sub>7</sub>   | 97                         | 0.3  | 8        | 2        | C <sub>7</sub> H <sub>13</sub>   |
| 56                       | 6.2   | 2        | 1        | C <sub>4</sub> H <sub>8</sub>   | 99                         | 0.4  | 0        | 0        |                                  |
| 57                       | 13.0  | 2        | 1        | C <sub>4</sub> H <sub>9</sub>   | 100                        | 1.0  | 0        | 0        |                                  |
| 58                       | 8.8   | 0        | 0        |                                 | 101                        | 19.7 | 6        | 2        | C <sub>6</sub> H <sub>13</sub> O |
| 59                       | 100.0 | 6        | 2        | C <sub>3</sub> H <sub>7</sub> O | 102                        | 1.4  | 0        | 0        |                                  |
| 60                       | 3.5   | 4        | 1        | C <sub>3</sub> H <sub>8</sub> O | 112                        | 2.2  | 4        | 1        | C <sub>8</sub> H <sub>16</sub>   |
| 69                       | 3.0   | 8        | 4        | C <sub>5</sub> H <sub>9</sub>   | 113                        | 0.3  | 1        | 1        | C <sub>8</sub> H <sub>17</sub>   |
| 70                       | 2.8   | 1        | 1        | C <sub>5</sub> H <sub>10</sub>  | 115                        | 0.0  | 8        | 2        | C <sub>7</sub> H <sub>15</sub> O |
| 71                       | 1.9   | 1        | 1        | C <sub>5</sub> H <sub>11</sub>  | 129                        | 0.4  | 0        | 0        |                                  |
| 72                       | 1.1   | 0        | 0        |                                 | 130                        | 0.1  | 1        | 1        | C <sub>8</sub> H <sub>18</sub> O |
| Number of Mechanisms: 83 |       |          |          |                                 | % not accounted for: 8.7 % |      |          |          |                                  |

TABLE 3.

Mass Spectrum of Dodecane-6-one (70 eV) /5/

| <i>m/e</i>                | %     | <i>M</i> | <i>F</i> | <i>Formula</i>                  | <i>m/e</i>                 | %    | <i>M</i> | <i>F</i> | <i>Formula</i>                    |
|---------------------------|-------|----------|----------|---------------------------------|----------------------------|------|----------|----------|-----------------------------------|
| 43                        | 100.0 | 6        | 2        | C <sub>2</sub> H <sub>3</sub> O | 97                         | 1.3  | 0        | 0        |                                   |
|                           |       | 1        | 1        | C <sub>3</sub> H <sub>7</sub>   | 99                         | 35.2 | 11       | 5        | C <sub>6</sub> H <sub>11</sub> O  |
| 44                        | 3.6   | 6        | 2        | C <sub>2</sub> H <sub>4</sub> O | 100                        | 2.6  | 1        | 1        | C <sub>6</sub> H <sub>12</sub> O  |
| 45                        | 1.3   | 6        | 4        | C <sub>2</sub> H <sub>5</sub> O | 110                        | 1.3  | 2        | 1        | C <sub>8</sub> H <sub>14</sub>    |
| 54                        | 1.1   | 3        | 2        | C <sub>4</sub> H <sub>6</sub>   | 111                        | <1   | 1        | 1        | C <sub>8</sub> H <sub>15</sub>    |
| 55                        | 17.6  | 0        | 0        |                                 | 112                        | <1   | 1        | 1        | C <sub>7</sub> H <sub>12</sub> O  |
| 56                        | 4.0   | 2        | 1        | C <sub>4</sub> H <sub>8</sub>   | 113                        | 31.2 | 11       | 6        | C <sub>7</sub> H <sub>13</sub> O  |
| 57                        | 10.7  | 1        | 1        | C <sub>3</sub> H <sub>5</sub> O | 114                        | 8.0  | 3        | 2        | C <sub>7</sub> H <sub>14</sub> O  |
|                           |       | 1        | 1        | C <sub>4</sub> H <sub>7</sub>   | 115                        | 2.7  | 1        | 1        | C <sub>7</sub> H <sub>15</sub> O  |
| 58                        | 48.5  | 6        | 1        | C <sub>3</sub> H <sub>7</sub> O | 124                        | 0.0  | 2        | 1        | C <sub>9</sub> H <sub>16</sub>    |
| 59                        | 12.3  | 5        | 3        | C <sub>3</sub> H <sub>6</sub> O | 125                        | <1   | 1        | 1        | C <sub>9</sub> H <sub>17</sub>    |
| 68                        | 3.9   | 0        | 0        |                                 | 126                        | <1   | 2        | 1        | C <sub>8</sub> H <sub>14</sub> O  |
| 69                        | 3.7   | 0        | 0        |                                 | 127                        | 4.8  | 9        | 2        | C <sub>8</sub> H <sub>15</sub> O  |
| 70                        | 2.9   | 2        | 1        | C <sub>5</sub> H <sub>10</sub>  | 128                        | 6.2  | 4        | 2        | C <sub>8</sub> H <sub>16</sub> O  |
| 71                        | 47.2  | 7        | 3        | C <sub>4</sub> H <sub>7</sub> O | 129                        | 1.6  | 0        | 0        |                                   |
|                           |       | 1        | 1        | C <sub>5</sub> H <sub>11</sub>  | 140                        | <1   | 1        | 1        | C <sub>9</sub> H <sub>16</sub> O  |
| 72                        | 7.3   | 0        | 0        |                                 | 141                        | 4.6  | 10       | 2        | C <sub>9</sub> H <sub>17</sub> O  |
| 81                        | 1.8   | 0        | 0        |                                 | 142                        | 1.0  | 1        | 1        | C <sub>9</sub> H <sub>18</sub> O  |
| 83                        | 2.2   | 0        | 0        |                                 | 152                        | 0.0  | 1        | 1        | C <sub>11</sub> H <sub>20</sub>   |
| 84                        | 1.1   | 0        | 0        |                                 | 154                        | <1   | 1        | 1        | C <sub>10</sub> H <sub>18</sub> O |
| 85                        | 17.4  | 9        | 3        | C <sub>5</sub> H <sub>9</sub> O | 155                        | 1.0  | 9        | 2        | C <sub>10</sub> H <sub>19</sub> O |
|                           |       | 1        | 1        | C <sub>6</sub> H <sub>13</sub>  | 167                        | <1   | 2        | 1        | C <sub>12</sub> H <sub>23</sub>   |
| 86                        | 3.8   | 0        | 0        |                                 | 169                        | <1   | 7        | 2        | C <sub>11</sub> H <sub>21</sub> O |
| 95                        | 2.0   | 0        | 0        |                                 | 184                        | 2.4  | 1        | 1        | C <sub>12</sub> H <sub>24</sub> O |
| 96                        | 1.1   | 0        | 0        |                                 |                            |      |          |          |                                   |
| Number of Mechanisms: 139 |       |          |          |                                 | % not accounted for: 12.0% |      |          |          |                                   |

An example of the type of output, found for the various mechanisms, is given in Figure 2<sup>3</sup> which illustrates the loss of water by the molecular ion of octane-3-ol, followed by the expulsion of a methyl radical to yield *m/e* 97 and finally *m/e* 41. Scheme 4 pictures, in the normal way, those mechanisms given in Figure 2. Water expulsion from the molecular ion is also proposed to proceed via four-membered and five-membered transition states.



SCHEME 4.

MECHANISM 1      MASS    130

ION 9

IONIC SPECIES

O(+.)H-CH--CH2-CH3CH2-CH2-CH2-CH3      C8H18O

9          3    2    1   4    5    6    7    8

MECHANISM 31      MASS    112

6MRTH6,9    HET9,3

IONIC SPECIES

C(+.)H--CH2-CH2-C(. )H-CH2-CH3CH2-CH3      C8H16

3          4    5    6          7    8    2    1

NEUTRAL SPECIES

OH2      H2O

9

MECHANISM 34      MASS    97

<sup>3</sup> The linear notation for chemical structures, used by the IG, has been explained previously /1,2,3/.

HOM7,8 MKB6,7  
 IONIC SPECIES  
C(+)H--CH2-CH3CH2-CH2-CH=CH2 C7H13  
 3 2 1 4 5 6 7

NEUTRAL SPECIES  
C(.)H3 CH3  
 8

---

MECHANISM 36 MASS 41

HET4,5 MKB2,3  
 IONIC SPECIES  
C(+)H2-CH=CH2 C3H5  
 5 6 7

NEUTRAL SPECIES  
CH-=CH2-CH3CH2 C4H8  
 3 2 1 4

---

FIGURE 2.

In Figure 3 is given the output from the treatment of dodecane-6-one by the IG, which results in  $m/e$  114, followed by  $m/e$  58, by two successive McLafferty rearrangements.

MECHANISM 1 MASS 184

ION 13  
 IONIC SPECIES  
O(+.)=C--CH2-CH2-CH2-CH2-CH2-CH3CH2-CH2-CH2-CH2-CH3 C12H24O  
 13 7 6 5 4 3 2 1 8 9 10 11 12

---

MECHANISM 3 MASS 114

6MRTH4,13 HOM5,6 MKB4,5 HOM7,3 MKB6,7  
 IONIC SPECIES  
O(+.)H-C=CH2-CH2-CH2-CH2-CH3CH2 C7H14O  
 13 7 8 9 10 11 12 6

NEUTRAL SPECIES  
CH3-CH2-CH2-CH=CH2 C5H10O  
 1 2 3 4 5

---

|                 |         |         |        |        |
|-----------------|---------|---------|--------|--------|
| MECHANISM       | 4       | MASS    | 58     |        |
| 6MRTH10,6       | HOM8,9  | MKB9,10 | HOM6,7 | MKB7,8 |
| IONIC SPECIES   |         |         |        |        |
| O(+.)H-C-       | =CH3CH2 |         |        | C3H6O  |
| 13              | 7       | 6       | 8      |        |
| NEUTRAL SPECIES |         |         |        |        |
| CH2=CH-CH2-CH3  |         |         |        | C4H8   |
| 9               | 10      | 11      | 12     |        |

---

FIGURE 3.

It can be seen from these varied examples that the set of primitive reaction steps, known to the ION GENERATOR program, and the heuristic conditions which guide their use, result in fragmentation schemes which a mass spectrometrist would normally postulate in the absence of deuterium labelling information. Moreover, as can be seen from the tables, the qualitative predictive ability of the IG is rather encouraging. In every case, all of the important ions are generated, and only few ions are proposed which are absent from the actual mass spectrum. Thus it can be said that the IG is a sound, first-order predictor of mass spectra.

At this point, one could envisage the achievement of a total mass spectral predictor. First, from preliminary experiments that have been performed with the IG on cyclic compounds, some new primitive reaction collections will need to be developed. After that, a model might eventually be found which could take the resultant reaction schemes that the IG produces currently, and would calculate therefrom a set of relative intensities.

#### Experimental Part

The IG program is implemented in FORTRAN on the CDC 3800 computer of the University of Geneva. Partial support of this work by a grant (2.478.71) of the Fonds National Suisse de la Recherche Scientifique is acknowledged. We thank Miss B. May for her technical assistance.

Bibliography

- /1/ A.B. Delfino and A. Buchs, Helv. 55, 2017(1972).
- /2/ A.B. Delfino and A. Buchs, Org. Mass Spectrom. 9, 459 (1974).
- /3/ A.B. Delfino and A. Buchs, Topics in Current Chemistry 39, 109-137 (1973), Springer-Verlag Berlin, Heidelberg, New York.
- /4/ Spectrum No. 466, Mass Spectrometry Data Centre, AWRE, Aldermaston, Berk.
- /5/ ibid, Spectrum No. 4300.

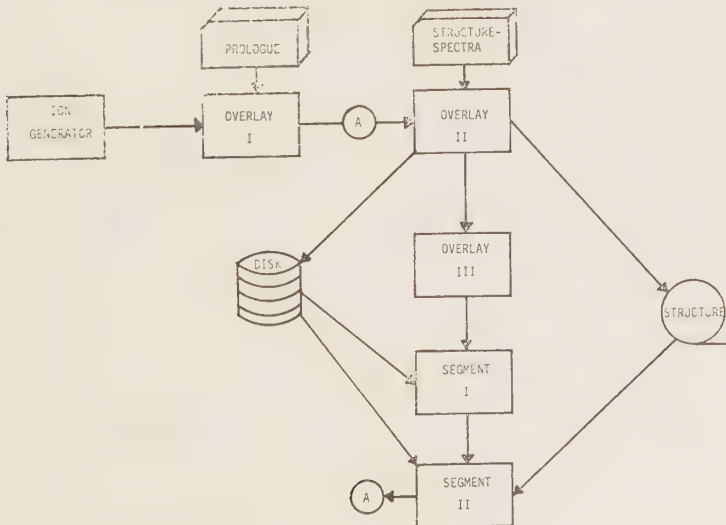
## Appendix F.

### PROGRAM DESCRIPTIONS OF THE ION GENERATOR SYSTEM

#### F.1 Introduction

This appendix is given to facilitate the comprehension of the program listing of the ION GENERATOR, available at the Mass Spectrometry Laboratory of the University of Geneva. With these short program descriptions or definitions, along with the comments interspersed throughout the program (there are approximately 1700 comments therein), one can quickly become familiar with the program logic.

The ION GENERATOR program (IGP) is implemented on and for the Control Data Corporation 3800. Were the IGP to be implemented on some other computer, some machine-dependent coding would have to be changed. No so-called "programming tricks" have been used, rather heavy use of information packing has been made to keep the size of the program reasonable. The IGP consists of a main program with three overlays. Overlay numbered three contains two segments in order to share a common block that need not be known to the other overlays. A very general flow-chart of this organization is given below.



The prologue data set, discussed in chapter V, is read by overlay numbered one, which is used but one time in any execution of the IGP. The variable number of structure-spectra pairs is read and processed by overlays numbered two and three. Those mechanisms generated by overlay two are stored in a direct access file for use by overlay three. The structure in virgin form is passed from overlay two to overlay three, segment two via a sequential file on the disk.

All of overlay one, and segment two of overlay three are kept as Macrosim /62/ files on tape volumes, and there is a card version of each one kept at the Mass Spectrometry Laboratory. In the following sections, a short description of every routine of this system will be given, grouped by overlay and segment.

#### *F.2 ION GENERATOR, The Main Program.*

The main program program of the IGP, entitled IONGENER, reserves all of the common blocks for data structures, initializes the random-access disk file, and calls the various overlays into play. Overlay numbered two, when it detects that no more structure-spectra pairs are left to be processed, calls the EXIT system program to terminate the execution of the IGP.

#### *F.3 Overlay I, Prologue Data Set Input.*

The prologue data set, described in chapter 5, is read and processed by overlay number one. The prologue data set includes format cards for the various dumps furnished by the IGP to diagnose errors, or to permit the user to follow the program through the prediction process on an example structure, along with the various options available to the user for program control. Also the prologue data set contains the ionization table and the periodic table. It also permits the user to suppress certain reactions during the run, to allow experiments to determine the efficaciousness of the various reactions with respect to the final results.

##### *F.3.1 Programs of Overlay I.*

MSEPTM1 - This program calls all of the subroutines to process the prologue

data set. It does no processing on its own.

- COMSET - This program sets all common blocks with various size constants, resets the ring memory, reads in the stack format cards, the options card, and the primitive reaction suppression card. It is in this program where the complete descriptions of the various common blocks are to be found.
- RSETRG - The routine which clears the ring memory system.
- PTREAD - This program reads the periodic table along with the hydrogen equivalents card. The periodic table can change from run to run. The order of the ion formulae is governed by the order of the atoms mentioned in the periodic table deck. It is then normal to have carbon as the first card, followed immediately by hydrogen, which is then followed by the rest of the atoms. There is space provided for 25 atoms.
- READIP - The ionization table is read by this routine. It also prepares tables which will be used by the ionization surveyor program, PDION, in overlay II.
- CLRALL - Variables, assumed to have certain values at the start of the run, are set to these values by this routine. It operates on three common blocks: the structure (STRUC), variables related to theory (THEORY), and the options (OPTIONS).

#### *F.4 Overlay II, The Mechanism Generator.*

The heart of the IGP is overlay two, because it does the mechanism generation process, while the other overlays exist only to provide a readable output form for the chemist-user. The structure-spectra pairs are read in at this time. After the program has accepted the structure and stored it in the common block (STRUC), a copy of this block as it was at that instant, is written into a temporary disk file to be read by overlay three, segment two. This is done, in case overlay two cannot finish the generation process, overlay three can still function on the partial results. The comparison spectrum can be null if it is not to be given to the program to annotate the mechanisms generated with the resulting mass' intensity from an actual spectrum.

Three tape volumes are used to store this overlay, Nos. CCI-2490,

CCI-2491 and CCI-2492. When program changes are made, updates of these files are made in "carousel" order: 2490 to 2491, 2491 to 2492, and 2492 to 2490. This provides protection against unforeseeable errors causing a total disaster. All recent changes are kept to be able to backup to a prior version by recreating the lost updates.

The programs, described in the next section, number 115, and there are 13 common blocks mentioned also. They are given in their order of appearance on the tapes.

#### *F.4.1 Programs of Overlay II.*

- DIARY - This common block contains the current mechanism. When fragmentation occurs, program PDDIAR puts a copy in the disk file for overlay III. When back-reactions are done, the needed information is retrieved from this block. The main matrix is used as a LIFO (Last In First Out) queue.
- DIARYARG - Any program that addresses a diary cell uses this dimension statement.
- IONIZE - The common block contains the ionization table and related items.
- LOGIC - The truth values are stored in a common block, because .TRUE. nor .FALSE. cannot be used directly as arguments. It contains the two forms, useable as arguments.
- OPTIONS - The optional information from the prologue data set, and related items are stored in this common block.
- PDSARGS - This dimension statement is needed by any program which addresses a surveyor stack cell.
- PERIOD - The periodic table, and all related informational items, are stored in this common block; this includes the ionic and neutral species masses of the current ionic structure, along with the ionic formula.
- PUSH - The common block contains the nodal stack space for surveyor programs. When push-down stacks are needed by working programs, such as COLRNG and DENDRAL to name two, they use the available space just after the last surveyor stack.
- RINGS - The ring memory data structure is stored in this common block.
- SPECTRA - This common block contains the comparison spectrum, read in by SPECIN.
- STRUC - The structure of the subject compound is stored in this block, along with the bond-locking matrices, radical and charge site vectors, and other related items.

- THEORY - This common block is used to store those variables and optional constants, which are related to the basic theory of the IGP.
- WEAK - Program WEAKON prepares these two tables of weak bonds and weakly held hydrogens. Any surveyor program who needs this information needs to have a copy of this common card set.

The above common block descriptions are followed by those of the programs.

- ACTION - All of the primitive reaction realization modules interface to this program, permitting easy inclusion of new reactions should they become necessary. All of the common chores of reaction realization are coded only once, and are done by this routine and a routine that it calls, PDDIAR.
- ADDNEW - This routine is used to inform the ring memory of a newly created single bond, which might have resulted in a ring formation.
- ADDRNG - A ring, found by program COLRNG, is stored in the ring memory by this program.
- ARNGNM - Given an atom number from the structure, this program returns all the ring names in which is found the given atom.
- BRBHAC - The 'heteroatom bond replacement by a hydrogen' is a name of a primitive reaction of the system. This reaction is realized on the current structure by this program.
- BRBHIN - The back-reaction module correspondent to BRBHAC is this program.
- BRBHPR - This program is the surveyor module correspondent to BRBHAC.
- CHERCH - When a comparison spectrum is given, this program is used to look up a low resolution mass' intensity. It is called from program ACTION.
- CHGADD - This memory system routine changes ring index addresses for deletions and during the ring space garbage collection sequence.
- CLRALL - The common blocks are reset to starting values when a new structure is about to be done by this routine.
- COLRNG - This routine finds all rings which have the atom given as an argument therein. Any rings which were not known before, are made known. All rings that were found are stored in the ring name table in block RINGS.
- COLSSA - This routine belongs to the CO loss mechanism set. It does the realization step.

- COLSSI - This routine is the back-reaction routine corresponding to COLSSA.
- COLSSP - This surveyor routine belongs to the CO loss primitive reaction.
- CTABIN - The structure is accepted and checked by this routine. It also assures that the ring memory knows of all the rings in the molecule.
- DBMVAB - This routine moves a double bond according to the needs of a double bond triggered mechanism's realization and back-reaction modules.
- DBNOMV - This logical function answers the question "Does the given atom of the structure make a double bond with any of its neighbors?".
- DENDRAL - The linear notation representation of a structure is formed by this program of the maze-following type. At the same time it recovers the formula of the substructure. A call to the program MASS right after a DENDRAL call will result in a mass calculation.
- DMPRNG - This routine prints out all of the variable from the ring memory common block in a highly readable format. The comment cards, describing the RING common block are in the DMPRNG deck.
- DOHETR - This is the realization module of the heterolysis of a heteroatom.
- DOHOMO - Regular  $\alpha$ -cleavage is realized by the application of this program.
- DOTRAN - This realization routine is used for the reaction entitled "Hydrogen rearrangement triggered by a radical site".
- EDGIRC - The ring index for the given ring edge is supplied by a call to this ring memory system program.
- EDGIST - A new ring index (name) is stored for the given ring edge by this ring memory program.
- ERNGNM - Given a table in which to put them, this program supplies a list of ring names which are the ones to which the given edge belongs. It also returns a count, which is zero if the given edge is not in a ring.
- EVLNAM - Given a ring name and the form of the ring desired (ordered or circuitual), this ring memory program returns the ring in the given form; hence the name of evaluation of a ring name.
- FALLNM - This routine returns the names of all the rings currently known to be in the structure or substructure.
- FLEXP - This routine answers the question, "Is the proposed ring or transition state flexible?". Since the answer depends strongly on the ring memory, this program should be considered as part of the ring memory system.

- FNDBFE - This routine returns all the single bonds that are  $\beta$  to the double bond given as (A,B) where the bonds attached to B are the  $\alpha$  bonds.
- FNDRNG - Since the FLEXP routine is given only the end points, but it needs the sequence of atoms making up the proposed ring or transition state, FNDRNG supplies the missing information.
- FORMUL - This routine uses the IHITS array, from the STRUC common, to calculate the formula. Therefore a maze-follower such as DENDRAL or THESEUS must have recently updated IHITS.
- FRNGNM - This ring memory routine is called by ERNGNM and ARNGNM to carry out their assigned duties with a single program.
- GCRNGI - When the available space of the ring index level of the ring memory is exhausted, an attempt to re-assemble any unused space resulting from deletion is made by this ring memory system program.
- GCRNGS - A garbage collection of deleted rings is done by GCRNGS. Afterwards, the rings still in use are in the lower memory, and the available space is in the higher memory starting at RNGPNT.
- GETEPD - This routine returns a table with all the heteroatoms in the given subset.
- GETMBD - This routine returns a table with the non-conjugated double bonds stored therein.
- HETRPR - A surveyor program; this routine corresponds to DOHETR.
- HEURFP - Unallowed primitive reaction sequences are detected by this routine and also if any steps have been suppressed via the suppression card of the prologue data set, the appropriate action is taken here.
- HLTSCR - This program dumps all common blocks in a highly readable format. It either stops thereafter or returns to the calling program depending on the value of the given logical variable.
- HOMOPR - This surveyor routine belongs to the  $\alpha$ -cleavage primitive reaction.
- HTRNGA - The primitive reaction, entitled "carbon skeleton rearrangement", uses this routine for reaction realizations.
- HTRNGI - This back-reaction unit corresponds to HTRNGA.
- HTRNGP - This routine is the surveyor unit which belongs to the same reaction as HTRNGA.
- INDRGO - Appropriate ring indexes will be created for the given ring by this ring memory program.
- INDRNG - All of the appropriate ring indexes, created via INDRGO, will result for the given ring when this program is called.

- INVERT - This switching routine interfaces with all of the back-reaction modules of the various chemical reaction sets and does all of the chores that would be common to all of them.
- INVHET - This back-reaction module corresponds to DOHETR.
- INVHOM - This back-reaction module corresponds to DOHOMO.
- INVION - The back-reaction of the ionization transformation is executed by this routine.
- INVTN - This back-reaction module corresponds to DOTRAN.
- IONSIT - An array, containing either the ionic site or the pair of indicators returned when the ionic site is part of a one-electron bond, is the output of this routine. If the one-electron bond is locked, then the routine returns a null set.
- ISRNGT - This routine answers the question, "Is this ring known already in the ring memory?". The ring must be given in canonical order.
- MASS - From a formula vector, this routine calculates the low-resolution mass and also the high-resolution value if that option has been taken.
- MKBACT - The reaction which forms bonds from two independent radical sites relies on this routine for the realization step.
- MKBINV - This back-reaction module correspond to MKBACT.
- MKBPRD - This surveyor unit corresponds to MKBACT.
- MLS1AC - The double bond triggered hydrogen rearrangement is broken into two independent steps. The first step which consists of the hydrogen shift with the double bond shift, but not including the single bond cleavage is done by the modules with the root name MLS1, of which this is the realization routine.
- MLS1IN - This is the back-reaction unit of the McLafferty first step primitive.
- MLS1PR - This is the surveyor module corresponding to MLS1AC.
- MLS2AC - The second step of the McLafferty (double bond triggered hydrogen rearrangement) consists of the single bond cleavage and any of the various satellite rearrangements which are possible if the original double bond had a radical site on the hydrogen receptor side. All programs in this set have the root name of MLS2. This routine is the realization routine.
- MLS2IN - This is the back-reaction module corresponding to MLS2AC.
- MLS2PR - This routine corresponds to the MLS2AC and does the surveyor part of the reaction set.

- MOEBAC - The reaction, that forms a one electron-bond from adjacent charge and radical sites, is realized by the use of this routine.
- MOEBIN - This routine does the back-reaction corresponding to the reaction done by MOEBAC.
- MOEBPR - This surveyor routine is that one corresponding to MOEBAC.
- MSEPTMA - The program which manages the whole mechanism generation process of overlay II is this routine. It is a main program, called in only by the overlay supervisor.
- NUMCON - The conversion of an integer to A1 format directly into a given position in a print buffer is the result of this program.
- ONEEVA - When an one-electron bond motivates a reaction, it must be thereafter dismantled, which is realized by this program unit.
- ONEEVI - This is the back-reaction program corresponding to ONEEVA.
- ONEEVP - This surveyor module corresponds to ONEEVA.
- ORDBND - This routine counts the number of bonds that a given atom has, including or excluding the hydrogens as an option.
- PDCELL - A program from the nodal stack system, this routine pushes down, on the current stack, a new cell.
- PDDIAR - This routine is the mechanism recording program. It accepts the diary cell record of the mechanism, assigns the mechanism number, and prepares and writes a record of the total mechanism on the direct access disk file.
- PDEOS - A program from the nodal stack system, it creates a stack by placing in the stack space a new end-of-stack indicator. This new stack becomes the current stack for PDCELL calls.
- PDEQUIV - This routine calls up in order all of the surveyor routines so that each can have a chance at the current nodal structure. It does not call the ionization surveyor, PDION, because ionization happens only at one time to the molecular structure.
- PDION - Surveying of the molecular structure for sites of ionization is done by this module once just after the structure is read in. This routine is called by MSEPTMA.
- PDSTRN - This routine prints out a buffer containing A1 stored characters; it is essentially an implementation of stream-type output.
- PUCELL - A program from the nodal stack system, this routine does an unstacking operation, finding either a cell or an end-of-stack indicator.

- PUDIAR - The topmost cell from the DIARY stack is retrieved, every time that an end-of-stack is sensed in the nodal stacks. With this cell, INVERT will call the appropriate back-reaction program to transform the structure to the structure from which it came.
- PUION - This realization routine goes to the ionization stack to get the information about the next ionization to be done, and then does it too.
- RCACTN - The ring closure reaction, triggered by a radical site, is realized by this program.
- RCINVR - This is the back-reaction routine corresponding to RCACTN.
- RCLTHA - The hydrogen push rearrangement is realized by this routine.
- RCLTHI - Back-reactions corresponding to RCLTHA are done by this module.
- RCLTHP - This is the surveyor unit for the reactions done by RCLTHA.
- RCPRD - This surveyor routine examines the structure for ring closure; this corresponds to RCACTN.
- RDARAC - The realizations of the retro-Diels-Alder reaction are accomplished by this program.
- RDARIN - This is the back-reaction module that corresponds to RDARAC.
- RDARPR - This surveyor module corresponds to RDARAC.
- ROVBAC - The homolytic relief of an overbonded heteroatom bearing the charge having a double bond is accomplished via this routine.
- ROVBIN - The back-reactions corresponding to ROVBAC are done by this program.
- ROVBPR - This is the surveyor module which corresponds to ROVBAC.
- ROWCRE - A routine from the ring memory system, this program aids in the flexibility algorithm by creating a row in the edge cross-reference matrix.
- RSETRG - A reset of the ring memory is accomplished by calling upon this module; this is essentially a "clear" of the memory.
- RSIONP - This routine examines at the behest of PDION, the rings in the molecule for the possibility of an ionization of a resonance ring.
- SAVRSTR - This routine saves or restores the IHITS and IPATHL arrays from the STRUC common, so that a maze-following routine can be called during a job where the results of a prior call are in use. The WILLOS program uses this routine so that HOMOPR will not be ruined by the call to WILLOS.
- SEARCH - This is the program that searches the structure to see if it has broken into two pieces as a result of the reaction that has just been realized thereon. If it finds a fragmentation, it recovers the charged species and the neutral species, measures their masses and calculates the metastable mass if required.

- SETDIF - This routine makes a set difference in order to be able to subtract unordered vectors from one another.
- SETNXP - A single reaction step can be specified for the next step in a mechanism by using this program. An example of its use is to assure that the second step of the McLafferty sequence follows after the first step, and no others.
- SHFTCR - This routine does a charge exchange given the old site and the new.
- SHFTCS - This program does the inverse of SHFTCR.
- SLSTAR - There is a vector in the DIARY common block that contains the last cell stored on the Diary stack. This program resets that array when the diary is empty at structure input time. Some theory variables are also clear at this time.
- SPECIN - This routine reads in the identification card and the comparison spectrum if there is one, or the count card punched with a one in the ion count field. Depending on the size of the spectrum, the ions are stored for a binary search or a random access search.
- SPHOMA - This is the realization routine for the homolysis with satellite rearrangement reaction. The surveyor program for this primitive is HOMOPR.
- SPHOMI - This routine does the back-reactions of the SPHOMA reaction.
- STBCHK - This routine checks for a stabilizing situation for the radical site if it were to end up there.
- STBCHP - This is the stability surveyor program that answers the question, "Is this a stable site for a radical at this time?".
- SUBOLD - A routine from the ring memory system, it directs the bond removal ring disturbance check and update process.
- THESEUS - This is a maze-follower which measures the least distances from a given point of origin to any other atom in the current subset of the molecule, up to a given greatest distance.
- TRNSPR - This is the surveyor routine which corresponds to DOTRAN.
- TRRADS - A vector, filled with the current radical sites of structure, is delivered to the calling program by this program.
- VUCCELL - A program in the nodal stack system, this routine gives the means for printing out all the stacks in the system without upsetting them.
- VUDIAR - This routine permits the contents of the diary stack without upsetting this stack.
- WEAKON - The weak bond table and the table containing the atom numbers of the sites where weakly held hydrogens can be found is created by this

program.

WIL00S - This routines predicts the effect on the mass of the proposed scission, before it is actually executed on the actual structure.

#### *F.5 Overlay III, Mechanism Sort and Print.*

Overlay III of the ION GENERATOR program is divided into two segments because the two segments use a large common block which is of no interest to the other overlays. This sort of program structure is demanded by that requirement. The main program of Overlay III is called 034SUP; it calls the two segments and does nothing else.

The first segment is just one program, SORTRP, which sorts the key variables of the mechanism records. These keys are the first 35 words of the 640 words of each record. The sorting algorithm is that of pointers, which avoids moving data around during the sort. The result of the sort is a disk file index which will be used by the second segment to print the mechanisms from low ionic mass to high, with ions of like formula grouped together. Any mass spectrally equivalent mechanisms are marked as such and are only mentioned and not totally repeated.

#### *F.5.1 Segment II, Overlay III, Program Descriptions.*

As was the case in Overlay II, segment II of Overlay III is stored as a MACROSIM file on tapes CCI-1380, CCI-1381 and CCI-1382. These programs are briefly described below. There are 15 common blocks and 37 programs.

CROSCM - This common block is used for the mechanism cross-reference system.  
The system enables a parent mechanism to mention all of its daughter ions even though it itself does not know which mechanisms they are.

DIARY - Same as Overlay II.

DIARYARG- Same as Overlay II.

IONIZE - Same as Overlay II.

LOGIC - Same as Overlay II.

MECIND - This common block contains the disk index prepared by segment I of this overlay.

OPTIONS - Same as Overlay II.

PERIOD - Same as Overlay II.

- PUSH - Same as Overlay II.
- RINGS - Same as Overlay II.
- SPECTRA - Same as Overlay II.
- STRUC - Same as Overlay II.
- THEORY - Same as Overlay II.
- WEAK - Same as Overlay II.
- ACTION - This is the same as ACTION in Overlay II, but with modifications to not only do the structural transformations but to print a report about it in full (ALLSTP option) or to use the normal restrained form.
- BRBHAC - This routine is similar to that of Overlay II, but it is modified to mechanism step transformations, reporting if called for, but it does not do the bond lock settings, ring memory calls, and other such chores that are done in Overlay II. All of the realization routines have been modified as such, and will be so indicated.
- COLSSA - Realization routine.
- CROSSR - This program collects the references for some future parent mechanism, when the daughter mechanism which know the parent ion from which they came, are processed.
- DBMVAB - Same as Overlay II.
- DENDRAL - This routine functions the same as that of Overlay II, but the print origin of the formula and structure is overshifted to column 20.
- DOHETR - Realization routine.
- DOHOMO - Realization routine.
- DOTRAN - Realization routine.
- DRVSTF - Sets the Diary viewing mechanism to do a FIFO pass at the stack, without actually upsetting the stack.
- FORMUL - Same as Overlay II.
- GATHER - This routine traverses the list structure of the cross-references to deliver a vector which can be used to print them on the final mechanism report.
- GETMEC - This program uses the disk file index to find the logically next mechanism record to fetch, it sets up the Diary stack with the information, and prints header and trailer lines of the report. It is this program that detects the equivalent mechanisms.
- HLTSCR - Same as Overlay II.
- HTRNGA - Realization routine.
- INTCRF - This is the program that clears the cross-reference system so that it can be reused by several structures.

MASS - Same as Overlay II.

MKBACT - Realization routine.

MLSTAC - Realization routine.

MLS2AC - Realization routine.

MOEBAC - Realization routine.

MSEPTMA - This is the main program, or entry point of the segment. It functions in much the same way as its counterpart of overlay II, except that it doesn't call up the surveyor programs, but does handle the print of the cross-reference list when the end of stack occurs in the Diary.

NUMCON - Same as Overlay II.

ONEEVA - Realization routine.

PDSTRN - This routine has been modified slightly to do the stream output with column 20 as the leftmost position on the page.

PUION - Realization routine.

RCACTN - Realization routine.

RCLTHA - Realization routine.

RDARAC - Realization routine.

ROVBAC - Realization routine.

RSSTRU - This routine is charged with the task of STRUC common refreshment at the end of each mechanism. To do this, it acquires a virgin copy of the common block from a temporary file. That copy was put there by CTABIN of Overlay II at the time that no changes had been made on it.

SAVRSTR - Same as Overlay II.

SEARCH - This routine had to be modified to avoid calculating values that are available from Overlay II calculations from the Diary cell itself.

SETDIF - Same as Overlay II.

SHFTCS - Same as Overlay II.

SPHOMA - Realization routine.

THESEUS - Same as Overlay II.

VDIARF - This is a FIFO viewer of the Diary stack.

## RESUME

Dans ce travail est décrit un programme heuristique, le ION GENERATOR, qui a été développé pour simuler la fragmentation que subissent les molécules des composés organiques quand elles sont soumises à l'impact d'électrons ionisants (70 eV) dans la source d'ions d'un spectromètre de masse. L'un des objectifs de ce travail était d'étudier jusqu'à quel point on pouvait simuler, par programmation, le processus intellectuel que suit le chimiste qui postule des mécanismes de fragmentation, pour expliquer la formation d'un spectre de masse. La difficulté essentielle qu'il a fallu surmonter a consisté à analyser minutieusement les concepts qu'utilise le chimiste, qui fait appel à son expérience, à sa connaissance de la technique des déplacements d'électrons (electron-bookkeeping theory), pour substituer à ces concepts un nombre fini d'actions élémentaires *programmables*. La nécessité de devoir définir de manière simple et rigoureuse un nombre restreint d'actions élémentaires par la combinaison desquelles il devenait possible de recréer la plupart des mécanismes de réaction, que l'on trouve dans les ouvrages classiques consacrés à la spectrométrie de masse, contribuera certainement beaucoup à l'élaboration d'une théorie des mécanismes de fragmentation, plus générale et mieux structurée qu'elle ne l'est actuellement. Les mécanismes que propose le ION GENERATOR sont construits par l'intervention successive, sur une structure ionique, des diverses opérations primitives, jusqu'à ce qu'une des transformations qui résulte de l'action d'une des opérations conduise à la formation d'un nouvel ion, par l'expulsion d'un fragment neutre. La séquence des opérations primitives qui conduit à une nouvelle espèce ionique, constitue le mécanisme de formation de ce nouvel ion.

Des résultats ont été publiés au fur et à mesure des progrès qui étaient réalisés dans le développement du programme. Dans un premier mémoire (Appendice B), les résultats obtenus pour la fragmentation de la 5-diéthylaminopentane-2-one sont présentés. Ce mémoire donne également une description détaillée de la manière de représenter n'importe quelle molécule organique non cyclique, par une notation linéaire. A ce moment seulement quatre opérations primitives étaient programmées au lieu des seize actuellement utilisées et qui sont décrites dans le chapitre IV.

Dans un second mémoire (Appendice A), un article de revue sur l'utilisation de l'informatique en spectrométrie de masse, un paragraphe est consacré aux développements intervenus après la publication du premier mémoire. Dans ce mémoire, l'accent est mis sur les mécanismes de réarrangement qui sont souvent à l'origine de la formation d'ions importants. Un composé relativement compliqué, le 2-n-hexyl-n-decanoate de méthyle a servi d'exemple pour examiner les performances du programme à ce stade de développement.

Dans un troisième mémoire (Appendice C), on décrit la fragmentation de molécules cycliques avec comme exemples la cyclohexanone, la 2-méthylpipéridine et la 2-méthylcyclohexanone. Le traitement de molécules cycliques a nécessité la mise au point d'une représentation linéaire des cycles, qui est expliquée dans cet article.

Au début, le domaine d'exploration du programme était limité à la création d'ions primaires, c'est-à-dire à la fragmentation d'ions moléculaires seulement. Cette limitation avait pour cause le manque de sélectivité dans l'action des opérations primitives, ce qui conduisait à la création d'un nombre de mécanismes beaucoup trop élevé. Tous les mécanismes proposés étaient certes corrects du point de vue de la théorie des déplacements d'électrons, mais aucune motivation réelle n'était à l'origine de la plupart d'entre eux. Une modification de la stratégie a alors été décidée, afin de permettre au programme de fragmenter non seulement des ions moléculaires, mais des ions de n'importe quelle génération. Cela a nécessité l'introduction de davantage de théorie, comme par exemple des notions sur la stabilité de certains arrangements structuraux, la résonance, etc..., afin de ne permettre l'action d'une opération élémentaire que si elle conduit à un arrangement atomique relativement favorable.

Les résultats atteints après cette amélioration sont décrits dans un manuscrit soumis pour publication (Appendice E). Dans ce manuscrit, des exemples de fragmentation sont donnés pour divers composés aliphatiques. On peut entre autres y relever que les spectres de masse qualitatifs ( $m/e$  créés, sans les intensités relatives des courants ioniques), sont en excellent accord avec les spectres expérimentaux. Des résultats analogues obtenus avec divers composés cycliques sont présentés et discutés dans le chapitre VI.

Le ION GENERATOR représente une approche originale et fructueuse vers un programme apte à prédire le spectre de masse d'un composé organique quelconque. L'intérêt de ce programme réside essentiellement dans sa conception. A aucun moment le ION GENERATOR n'a besoin de reconnaître à quelle classe de substances une molécule appartient pour la fragmenter. Il n'y a pas de mécanisme préprogrammé, mais un jeu d'actions élémentaires avec lequel les mécanismes sont construits chaque fois *de novo*, en fonction de l'environnement moléculaire traité.





